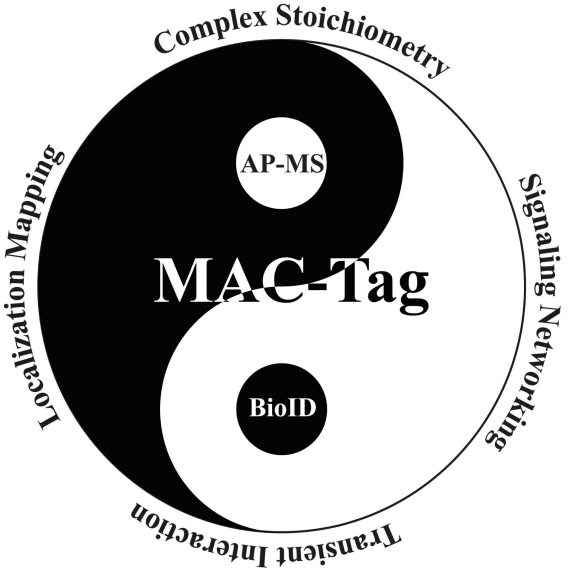
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

XIAONAN LIU

A NOVEL MULTIPLEXED INTERACTOME PROTEOMICS APPROACH FOR STUDYING CELLULAR SIGNALING



INSTITUTE OF BIOTECHNOLOGY
HELSINKI INSTITUTE OF LIFE SCIENCE HiLIFE AND
DIVISION OF GENETICS
DEPARTMENT OF BIOSCIENCES
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
DOCTORAL PROGRAMME IN INTEGRATIVE LIFE SCIENCE
UNIVERSITY OF HELSINKI

A novel multiplexed interactome proteomics approach for studying cellular signaling

Xiaonan Liu

Institute of Biotechnology

Helsinki Institute of Life Science

and

Division of Genetics

Department of Biosciences

Faculty of Biological and Environmental Sciences

and

Doctoral Program in Integrative Life Science

University of Helsinki

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki, in Metsätalo, Room 1, Unioninkatu 40, on the 8th of May, 2020 at 2 PM

Supervisor Docent Markku Varjosalo, Ph.D.
Group leader, Director of Proteomics Unit
Institute of Biotechnology
University of Helsinki, Helsinki, Finland

Thesis Committee Academy Professor Pekka Lappalainen, Ph.D.
Institute of Biotechnology
University of Helsinki
Finland

Professor Juha Partanen, Ph.D.
Faculty of Biological and Environmental Sciences
University of Helsinki
Finland

Reviewers Adjunct Professor Pieta Mattila, Ph.D.
Institute of Biomedicine, Turku Bioscience
University of Turku
Finland

Associate Professor Kirsi Rilla, Ph.D.
School of Medicine, Institute of Biomedicine
University of Eastern Finland
Finland

Opponent Franz Herzog, Ph.D.
Group Leader, Gene Center,
Department of Chemistry and Biochemistry
Ludwig Maximilian University of Munich
Germany

Custodian Professor Ville Hietakangas, Ph.D.
Centre of Excellence in Stem Cell Metabolism
Faculty of Biological and Environmental Sciences
University of Helsinki
Finland

Dissertationes scholae doctoralis ad sanitatem investigandam Universitatis Helsinkiensis
ISBN 978-951-51-6010-2 (paperback)
ISBN 978-951-51-6011-9 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)
https://ethesis.helsinki.fi/
Cover layout by Anita Tienhaara
Painosalama Oy, Turku, Finland 2020

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis contains material from the following original articles that referred to in the text by their Roman numerals I-IV. The rights have been granted by publishers to include the material in dissertation.

I. **Liu, X**., K. Salokas, F. Tamene, Y. Jiu, R. G. Weldatsadik, T. Ohman and M. Varjosalo (2018). "An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations." Nat Commun 9(1): 1188.
   - XL did the cloning, generated all the cell lines, performed both AP-MS and BioID experiments, and data analysis for the study. XL participated in designing the study, performed experimental work, data analysis, interpreted the results, and wrote the manuscript.

II. Konovalova, S., **X. Liu**, P. Manjunath, S. Baral, N. Neupane, T. Hilander, Y. Yang, D. Balboa, M. Terzioglu, L. Euro, M. Varjosalo and H. Tyynismaa (2018). "Redox regulation of GRPEL2 nucleotide exchange factor for mitochondrial HSP70 chaperone." Redox Biol 19: 37-45.
   - XL performed BioID experiments, data analysis and contributed to writing the corresponding section of the manuscript.

III. **X. Liu**, K. Salokas, R. G. Weldatsadik, L. Gawriyski, M. Varjosalo. Analyzing protein interactions by MAC-tag approach and MS-microscopy system. (Unpublished Manuscript, 2019).
   - XL participated in designing the study, performed experimental work, data analysis, interpreted the results, and wrote the manuscript.

IV. **X. Liu**, K. Salokas, K. Demir, P. Beachy, M. Varjosalo. A dynamic interaction landscape of the mammalian Hh signaling pathway. (Unpublished Manuscript, 2019).
   - XL participated in designing the study, performed experimental work, data analysis, interpreted the results, and wrote the manuscript.

ABBREVIATIONS

AP	Affinity purification
APEX	Engineered ascorbate peroxidase
ARL13B	ADP Ribosylation Factor Like GTPase 13B
ARRB	β-Arrestin
BAC	Bacterial artificial chromosome
Bbsome	Bardet-Biedl syndrome (BBS) proteins
Bcl-2	B-cell lymphoma 2
BCR	B-cell receptor
BET1	Golgi vesicular membrane-trafficking protein p18
BIFC	Bimolecular Fluorescence Complementation
BioID	Proximity-dependent biotin identification
BOC	Brother of CDO
CCDC85C	Coiled-coil domain-containing protein 85C
CDK	Cyclin-dependent kinase
CDO	CAM-related/down-regulated by oncogenes
CHL1	Close homolog of L1
Ci	Cubitus interruptus
CK1	Casein kinase 1
CO-IP	Co-immunoprecipitation
CRAPome	Contaminant repository for affinity purification-mass spectrometry data
CRD	Conserved cystein-rich domain
CTD	C-terminal domain
DHh	Desert Hedgehog
DISP	Dispatched
DLBCL	Diffuse large B-cell lymphoma cells
DLG5	Disks large homolog 5
DRAL	Down-regulated in rhabdomyosarcoma LIM domain protein
DSSO	Disuccinimidyl sulfoxide
DYNC2H1	Dynein Cytoplasmic 2 Heavy Chain 1
ECD	Extracellular domains
EFCAB7	EF-hand calcium-binding domain-containing protein 7
EM	Electron Microscopy
ER	Endoplasmic reticulum
EVC	Ellis-van Creveld syndrome protein
EVC2	Limbin
FACS	Fluorescence activated cell sorting
FBXl17	F-box and leucine-rich repeat protein 17

FDA	Food and Drug Administration
FDR	False discovery rate
FKBP	FK506-binding protein
FLAG	FLAG peptide tag
FRB	FKBP-rapamycin-binding domain
FRET	Fluorescence resonance energy transfer
G proteins	Guanine nucleotide-binding protein
GAS1	Growth arrest-specific 1
GFP	Aequorea Victoria Green fluorescent protein
GLI	Glioma-associated oncogene homolog
GLI1	Zinc finger protein GLI1
GLI2	Zinc finger protein GLI2
GLI3	Transcriptional activator GLI3
GLI3R	Repressor forms of GLI3
GO	Gene ontology
GOI	Gene of interest
GPR161	G Protein-Coupled Receptor 161
GSK3β	Glycogen synthase kinase-3 beta
HA	Human influenza hemagglutinin peptide tag
HCI	High-confidence interaction
HCIP	High-confidence candidate interacting protein
HH	Hedgehog
HHIP	Hedgehog interacting proteins
HSPGs	Heparan sulfate proteoglycans
IDR	Intrinsically disordered region
IFT	Intraflagellar transport
IHH	Indian Hedgehog
ILK	Integrin-linked kinase
IMM	Inner mitochondrial membrane
IMS	Intermembrane space
IQCE	IQ-domain containing protein E
KIF7	Kinesin family 7
LAP2β	Lamina-associated polypeptide 2β
LC	Liquid chromatography
LD	Linker domain
MAC-tag	Multiple Approaches Combined-tag system
MBP	Maltose binding protein
MEFs	Mouse embryonic fibroblast cell
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry

NEDD	Neural precursor cell expressed developmentally down-regulated protein
NEK2A	NIMA-related expressed kinase 2A
NPCs	Nuclear pore complexes
OMM	Outer mitochondrial membrane
PKA	cAMP-dependent protein kinase A
POI	Protein of interest
PPI	Protein-protein interaction
PTCH1	Patched1
PTM	Post-translational modification
QUBIC	Quantitative BAC-GFP interactomics
SAINT	Significance analysis of interactomes
SAP18	Sin3-associated polypeptide 18
SCUBE2	Scube 2
SHH	Sonic Hedgehog
SLTM	SAFB-like transcription modulator
SMO	Smoothened
SMURF	Smad ubiquitin regulatory factor
SPOP	Speckle-type POZ protein
SSD	Sterol-sensing domain
SUFU	Suppressor of Fused
TAD	Trans-activating domain
TGFβ	Transforming Growth Factor β
TM	Transmembrane domain
TUCAN	Tumor up-regulated CARD-containing antagonist of caspase-9
TULP3	Tubby-like protein 3
USP8	Ubiquitin-specific protease 8
VDAC	Voltage-dependent anion channel
VPS34	Vacuolar protein sorting 34
XIAP	X chromosome-linked IAP
XL-MS	Crosslinking mass spectrometry
YAP1	Yes-associated protein 1

ABSTRACT

Cellular compartment organization, signaling transduction, and regulatory processes are governed by protein-protein interactions (PPIs). The affinity purification coupled mass spectrometry (AP-MS) and proximity-labeling methods such as BioID have been wildly applied to understand protein interaction networks. AP-MS methods are suited to identify and quantify protein interaction and protein complex stoichiometries, while the BioID method provides information regarding transient or close-proximity interactions. Both AP-MS and BioID can identify distinct interactions, and complementary provides a more comprehensive view of a protein's interactome. In this study, we developed an integrated approach utilizing 'multiple approaches combined' (MAC)-tag that utilizes both AP-MS and BioID in a single construct to parallel explore the protein interactome.

First, we systematically applied the MAC-tag approach to 18 *bona fide* localization markers to generate a molecular context database for every cellular compartment. Using this molecular context as the reference, alignment interaction profile of other proteins of interest (POIs) on it can reveal the subcellular localization of the POIs. We named this reference database, mass spectrometry (MS)-microscopy. Similarly, we applied the MAC-tag approach on three sub-organelle localization markers of mitochondria to set up a sub-organelle molecular context proteome map. Using this sub-organelle reference proteome map, we could detect sub-mitochondrial localization of POIs by MS-microscopy system. Furthermore, by comparing interaction data from several MAC-tagged mitochondrial proteins, we could define the unique interactors of GrpE-like proteins (GRPEL1 and GRPEL2) to investigate their specific roles in mammalian mitochondria.

Next, we utilized the MAC-tag approach to map the interactome of the mammalian Hedgehog signaling. We monitored the dynamic PPI network surrounding the core components in the absence and presence of Hh ligand. Our comparative PPI analysis showed that the interaction network recapitulates many known ciliary proteins, asserting the essential role of primary cilia in Hh signaling. Meanwhile, we characterized the interaction network at the level of cilium progression by monitoring the PPIs under cilia absence, ciliogenesis, and cilia absorption conditions. The significant overlap of the interaction between three states indicated that proteins could regulate the Hh signaling outside the primary cilia. Additionally, we expanded the coverage of MS-microscopy to the primary cilium, a transient organelle of the cell. Although what we present is just a glimpse of the Hh interaction network under ligand stimulation or specific growth conditions, these PPIs will extend our knowledge on the mammalian Hh signaling and help us to development of novel therapeutic strategies to inhibit abnormal Hh signaling in cancer.

Collectively, the presented thesis introduces the MAC-tag approach that can be used for mapping local protein interactome, predicting subcellular localization of POIs, and reconstructing the interaction network of the signaling pathway.

I LITERATURE REVIEW

1 Interactome Proteomics

Protein-protein interactions (PPIs) play a fundamental role in the regulation of the biological processes, and the development of the diseases. The classical definition of PPI focuses on the physical contacts of two or more protein molecular [1, 2]. It can be classified into stable or transient interaction. Stable PPIs often form a functional complex, which is not easily disrupted over a long time course, while transient PPIs occur only briefly as part of a biological process and are difficult to detect [3]. Most biological events require the coordination of both types of interactions.

An expanded view defines the PPI in the context of its network of interactions [4, 5]. Each protein interacts with a number of partners that also interact with other proteins stably or transiently. All these interactions form an interaction network. Therefore, an interactome refers to the complete network of protein interactions in a particular cell [6]. The interactome proteomics studies proteome-wide protein associations and is an effective way to understand the biological processes inside the cell.

1.1 Liquid Chromatography Mass Spectrometry (LC-MS)

Liquid chromatography (LC) coupled with mass spectrometry (MS) (LC-MS) is an analytical technique that combines the physical and chemical separation capabilities of LC with the mass spectrum detection capabilities of MS.

The LC-MS system allows for the online separation of complex samples and has developed as a major analytical platform for high-throughput proteomics studies. Various LC-MS techniques have been developed to quantify and qualify proteins on both intact-protein and peptide levels [7, 8]. However, the majority of LC-MS based approaches are performed at peptide levels for the sensitivity, resolution, and fast execution of MS. Bottom-up protein identification is the most widely used approach [9]. It requires several steps of sample preparation, including cell lysis, protein separation, and proteolytic digestion. Once these steps are complete, the resultant peptides are further separated by LC according to their hydrophobicity. Eluted peptides from the LC can be ionized by electrospray ionization and analyzed by MS [10].

Once inside the mass spectrometer, multiple-charged peptide ions are separated and resolved in mass analyzer according to their mass to charge (m/z) ratio. After the first MS mass scan, ions with given m/z are selected undergo collision induced-dissociation (CID) with inert gas, such as argon or helium, to produce daughter ions [11]. The m/z of the daughter ions will be scanned and recorded to produce tandem mass spectrums (MS/MS) for identification (**Figure 1**). Identification of fragmentation spectra as peptide sequences is performed by computer algorithms such as Mascot [12], SEQUEST [13], and MaxQuant/Andromeda [14]. Protein identification is based on the detection of peptides.

Two distinct peptides from the same protein were the minimum requirement for reliable identification of its presence in the original complex sample [15].

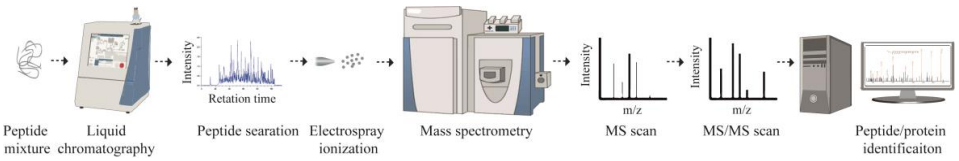


Figure 1. MS workflow. The peptides mixture is loaded on a LC column for liquid chromatography followed by electrospray ionization. Ions are then subjected to MS scan. The top N most abundant peptides are selected for CID and MS/MS scans. The acquired tandem mass spectrums are recorded and used to identify peptides and associated proteins (modified).

1.2 Affinity purification (AP)

The characterization of PPIs requires the isolation of protein complexes close to their physiological states. The affinity-purification (AP) experiment is probably one of the most frequently employed methods to purify protein complexes [16, 17]. In a typical experiment, a protein of interest (POI)/bait protein is expressed with an epitope tag in the cultured cell. The cell lysate is obtained by mild/ non-denaturing lysis conditions to preserve the protein complexes. The immobilized antibody that specifically recognizes the epitope tag can capture the protein complexes from cell lysate. Affinity-purification allows studying proteins under their molecular context. With the LC-MS system, thousands of peptides/proteins can be identified in a single affinity-purification analysis. Thus, affinity-purification coupled to mass spectrometry (AP-MS) has become the method of choice to discover the novel interacting partners [17, 18].

The epitope tag used for affinity-purification could be a protein or peptide (peptides or proteins). It can be fused either to the N- terminus, C-terminus, or even into the middle of the protein sequence. The fused tag positions should not affect the signal sequences, post-translation modification sites, or any regions that can comprise the function of the POI. In this chapter, we will review the advantages and disadvantages of several affinity tags that are frequently used in mammalian systems.

1.2.1 Protein tags

Several proteins could be used as affinity tags, such as *Aequorea Victoria* green fluorescent protein (GFP) [19], and maltose binding protein (MBP) [20].

GFP offers a specific advantage since it has autofluorescence when exposed to UV light. Thus, fluorescence-activated cell sorting (FACS) can be applied to isolate transgene-positive cells from negative clones [21]. Moreover, subcellular localization of the bait protein can be observed using microscopy. GFP-tagged proteins can be enriched by immobilized GFP antibody and dissociated from the antibody by lowering the pH.

Hubner and colleagues [22] tagged genes of interest (GOIs) with GFP using bacterial artificial chromosome (BAC) transgenes to generate the stable transfected HEK293 cell lines that express POIs at endogenous level. The purification step was performed on the automated liquid-handling platform with the anti-GFP antibody coupled with magnetic beads. After enrichment and on beads digestion, the peptide sample was analyzed by MS. The protein interaction data were quantified to detect the specific interactions from the background. This technique was named as quantitative BAC-GFP interactomics (QUBIC). When referring to the quantification of protein complexes, QUBIC performing in label-free format can be as efficient as it combining with traditional isotope labeling (SILAC).

Later, the QUBIC method was adopted to generate a library of 1,125 stable BAC-GFP HeLa cell lines to globally map the human interactome [23]. These PPIs (28,000 interactions) assemble a large-scale map of the human interactome, which can be used as a reference for other biological studies. Furthermore, the result provides a unique perspective on the detection of specific interaction, estimation of interaction stoichiometry, and measurement of cellular abundances of the interactor, revealing that weak, substoichiometric interactions dominate the PPI network.

MBP is a ~42 kDa protein of *Escherichia coli* (*E. coli*), which is responsible for the uptake, breakdown, and transport of maltodextrin. Although the exact mechanism of protein solubility enhancement by the MBP-tag remains uncertain [24], it can stabilize the target protein and improve the solubility of the fused protein. MBP is one of the most frequently used protein expression tags for crystallization [25]. It expresses in mammalian cells in monomeric form and can be applied to secreted proteins. The MBP-tagged protein binds to amylose resin-based chromatography and can be released by neutralizing pH with maltose-containing buffer conditions. However, MBP fused proteins do not always bind efficiently to the amylose resin [26].

1.2.2 **Peptide tags**

The size of the affinity tag is important since the tag may interfere with the biological functions of the tagged protein, such as protein folding, or subcellular localization. Therefore, small-size tags are commonly favoured in affinity purification. A series of short peptide-based tags have been developed, such as HA tag [27], c-Myc tag [28], FLAG-tag [29] and Strep-tag [30]. All these tags can be recognized by their corresponding antibodies.

The FLAG-tag is a hydrophilic octapeptide (DYKDDDDK) that can be recognized by specific antibodies [29]. There are monoclonal anti-FLAG antibodies, including M1, M2, and M5 that can bind the FLAG epitope [31]. Each antibody exhibits different recognition and binding characteristics. Polyclonal antibodies with higher affinity than these monoclonals to detect FLAG-tag are also available from few companies. The elution of FLAG-tagged proteins is performed with FLAG peptide or low pH glycine buffer. Additionally, the FLAG-tag has been used endogenously to tag the small (eS17) and large

(eL36) ribosomal subunits in mouse embryonic stem cells (ESCs) [32] to identify ribosomal associated proteins (RAPs).

Similarly, the Strep-tag is an octapeptide (WSHPQFEK or AWAHPQPGG) that binds to streptavidin [30]. The Strep-tag derivative, twin-Strep-tag (StrepIII-tag) contains two copies of Strep-tag [33]. With engineered streptavidin, Strep-Tactin®, the binding affinity with StrepIII-tag can increase 100 times higher. The high binding affinity enables one-step purification of the fused Avidin can be used to pre-wash the matrix before sample loading to prevent endogenous biotinylated proteins binding during purification [34].

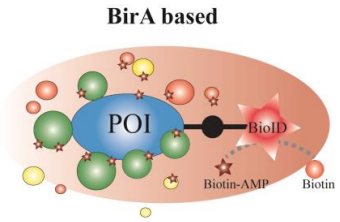
The tandem affinity purification (TAP) combining two affinity tags has been developed [35] to reduce the chance of contaminants retained in the eluate. The TAP strategy contains two affinity purification steps. The application of the second step removes contaminants that might be specific for the first purification to achieve higher purity of the sample. For example, the TAP-tag system containing a StrepIII-tag and HA-tag was used to determine the protein interaction networks of the 57 CMGC kinases [36]. The stable HEK293 cell lines were generated for each of the 57 kinases, and kinase complexes were purified by a two-step sequential affinity purification. The cell lysate was first loaded on Strep-Tactin chromatography. The elution was then incubated with anti-HA agarose to obtain a higher purity of target protein complexes for MS analysis. They identified 652 high-confidence kinase-protein interactions and revealed a broad set of interactions involving in controlling human disease pathways.

1.3 **Proximity-based labeling (PL)**

AP-MS is the choice to detect stable interactions. However, many interactions are transient and occur only in specific cellular compartments or at particular stages of development [37]. It is ineffective to apply AP-MS to study these dynamic interactions. Thus, several alternative approaches based on proximity-detection have been developed in recent years as powerful complementary approaches to AP-MS. Here we focus on applications and limitations of several proximity labeling-based methods coupled with mass spectrometry.

1.3.1 **BioID and BioID2**

The BioID [38] is based on a mutant form of biotin protein ligase (BPLs, *Escherichia coli* BirA R118G, BirA*) [39] fused with a POI. The BirA mutant produces highly reactive biotinoyl-5′-AMP that can diffuse to biotinylate proteins (**Figure 2**) that are interacting directly or indirectly with the bait protein within a radius of ~10 nm [40]. The covalent addition of biotin allows the labeled proteins to be captured in a single step streptavidin purification.



BioID; BioID2; TurboID; miniTurboID

Figure 2. Schematic of the BirA based proximity-labeling method. A mutated version of BirA is fused with POI. It can utilize biotin and ATP to generate reactive biotin-AMP that can label proximal proteins.

BioID has made significant contributions to proteomics studies since 2012. It has been applied for PPI studies of insoluble, inaccessible, or low-abundance structures, including chromatin, centrosomes, cell junctions, signaling pathways, HIV-host cells [41]. Moreover, BioID has been used to study protein interactions *in vivo* (iBioID) [42] and protein-RAN interactions (RaPID) [43].

Despite the widespread application of BioID, the major disadvantage of BioID is the size of BirA* (~35 kDa) that could potentially affect the localization and/or function of the POI. For example, BirA* tagging can restrict lamina-associated polypeptide 2β (LAP2β) passage through nuclear pore complexes (NPCs) [44]. Additionally, BioID has very slow kinetics, which necessitates labeling with biotin for 16-18 hours and even longer for *in vivo* labeling to produce sufficient biotinylated material for proteomic analysis. Furthermore, since the efficacy of the biotinylation depends on the number and availability of lysine in proteins (Roux et al., 2013), the abundance of the purified biotinylated proteins does not necessarily correlate with the strength or stoichiometry of the association.

The BioID2 [45] is the humanized biotin ligase from *Aquifex aeolicus* with a mutation (R40G, ~27 kDa) in the biotin catalytic domain (**Figure 2**). It functions as BioID but is much smaller. Moreover, BioID2 has a higher biotin affinity than BioID and requires less biotin for efficient labeling. The biotinylation range of BioID2 can be increased by adding a flexible linker (13 repeats of GGGGS) between tag and POI. Thus, BioID2 can be used to study protein complexes with larger spatial dimensions that exceed the biotinylation range of BioID. The caveat of BioID2 approach is its strong biotinylation ability using endogenous biotin, which prevents precise temporal control over labeling.

1.3.2 **TurboID and miniTurboID**

TurboID and miniTurboID (a truncated version of TurboID) are the latest proximity labeling techniques [46] that combine the simplicity and non-toxicity of BioID with higher biotinylation efficiency.

TurboID is a 35 kDa biotin protein ligase, with 15 mutations relative to wild-type BirA. MiniTurbo is a 28 kDa protein with the N-terminal domain deletion and 13 mutations relative to wild-type BirA (*Escherichia coli*). Therefore, they work similarly as BioID (**Figure 2**). Among all the mutated versions of BirA (BirA*), TurboID is the most active in biotin labeling in shorter time windows (10 min-1h), and miniTurboID is 1.5- to two-fold less active than TurboID. Both of them are 7-26 fold more active than BioID, enabling proximity labeling in just 10 minutes with little difference in proteome outputs when compared with the results from BioID for 18 hours labeling. Unlike BioID, which is mainly effective at 37°C, TurboID and miniTurboID are active in a wide range of temperature and can be easily applied on *drosophila* (grow at 25°C) and *c.elegans* (grow at 20°C) [46].

TurboID has limitations that should be considered during experimental design. It is very active and can use the endogenous biotin when lacking the addition of exogenous biotin, especially for the application *in vivo*. Flies ubiquitously expressing TurboID without biotin supplementation showed a smaller size and decreased survival [46]. Since TurboID consumes all the endogenous biotin, it results in a cellular biotin deficiency condition. MiniTurboID can be used as an alternative method to avoid such situations. However, miniTurbo exhibits less stable than TurboID due to the removal of its N-terminal domain, leading to lower expression levels in cells [46].

1.3.3 **2C-BioID**

2C-BioID, is established on the application of the FK506-binding protein (FKBP) and FKBP-rapamycin-binding domain (FRB) proteins that do not interact without rapamycin but form a tight ternary complex in its presence [47]. In this system, the BioID and POI are kept separately to fuse with FRB and FKBP (**Figure 3**).

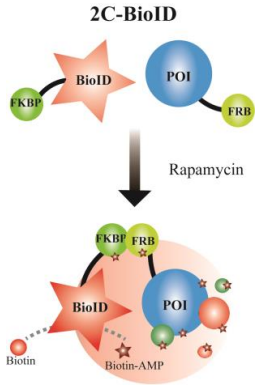


Figure 3. Schematic of the 2C-BioID method. BioID is fused with FKBP, while the POI is fused with FRB. In the presence of rapamycin, BioID and POI are brought into proximity. The biotinylation will be initiated by exogenous biotin.

Compared to BioID, 2C-BioID overcomes the limitation of that size of BioID/BioID2 that could affect the localization and/or function of the POI.

1.3.4 **Split-BioID**

The split-BioID assay is based on the expression of two halves of non-functional fragments of BioID with two POIs. The two pieces of BioID can reassemble and restore the biotinylation activity on the interaction of two POIs (**Figure 4**). BirA (*Escherichia coli* BirA R118G) can be split at multiple sites (sites Q140/E141 or E256/G257) [48, 49]. However, comparative examination suggests that the E256/G257 split regain a higher ligase activity [49].

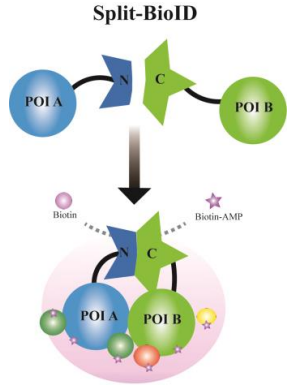


Figure 4. Schematic of the Split-BioID method. Two inactive fragments of BioID are fused to two putative interacting POI-A and POI-B. Upon interaction of POIs, BioID activity is restored and mediates the biotinylation of proximal proteins.

Split-BioID was designed mainly to validate binary interactions rather than for a high-throughput assay. Compared to BioID, split-BioID generates much less background biotinylation because only proteins that assemble around this pair of interacting proteins are labeled.

1.3.5 Engineered Ascorbate PEroXidase (APEX) and APEX2

APEX is a ~28 kDa engineered ascorbate peroxidase derived from pea or soybean [50, 51]. APEX uses hydrogen peroxide (H_2O_2) as an oxidant to convert biotin-phenol into biotin-phenoxyl radical (**Figure 5**), which has a half-life of <1 ms and can label electron-rich amino acid residues (Tyr, Trp, His, and Cys) of proteins in the vicinity of APEX [50]. The labeling radius of APEX is ~20 nm [51].

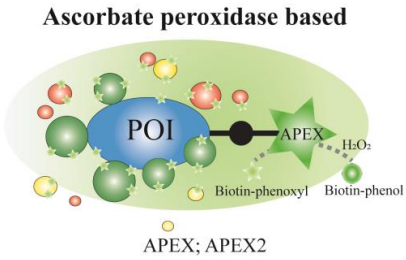


Figure 5. Schematic of ascorbate peroxidase based proximity-labeling method. A POI is fused to a mutant version of Ascorbate peroxide (APEX/APEX2). Upon on biotin-phenol and H_2O_2 incubation, APEX/APEX2 produces reactive biotin–phenoxyl molecules to label proximal proteins.

APEX requires less than one minute to label proximal proteins rather than the 16-24 hours required for the BioID. The high activate feature of APEX makes it suitable to study time-resolved PPIs and transient PPIs in a fast signaling turnover. A disadvantage of APEX is that biotinylation requires hydrogen peroxide treatment of cells, which potentially could affect cell oxidative status and cause cellular stress.

Comparing with APEX, APEX2 [52] works the same way (**Figure 5**). However, APEX2 has greatly improved cellular activity and sensitivity for proteomic tagging and EM labeling. Additionally, APEX2 is much more resistant to H_2O_2-induced inhibition than APEX [52].

1.3.6 Split-APEX2

A split approach of APEX2 can be achieved by splitting into a 200-amino-acid N-terminal "AP" fragment possessing of nine mutations (K22R, R24G, G50R, K61R, H62Y, I165L,

N72S, P125L, I185V) relative to APEX2 and a 50-amino-acid C-terminal "EX" fragment [53]. These split fragments are inactive on their own but reconstitute to give peroxidase activity when driven together by the protein-protein interaction (**Figure 6**).

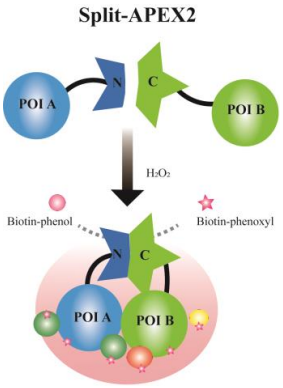


Figure 6. Schematic of the Split-APEX2 method. Two inactive fragments of APEX2 are fused to two putative interacting POI-A and POI-B. Upon interaction of POIs, APEX2 activity is restored and produces active biotin–phenoxyl molecules to label proximal proteins.

To characterize the contact sites between the mitochondrial outer membrane and the ER membrane, author fused AP with FKBP that targeted to the outer mitochondrial membrane (OMM), and EX with FRB that targeted to the endoplasmic reticulum (ER) membrane in COS7 cells. When cells were treated with rapamycin, FKBP interacts with FRB and split-APEX2 reconstitute to APEX2 to biotinylate nearby proteins. The biotinylation radius of recombined APEX2 is ~25nm [53].

1.3.7 Selective Proteomic Proximity Labeling Assay using Tyramide (SPPLAT)

Selective proteomic proximity labeling assay using tyramide (SPPLAT) [54] utilizes the reaction between horseradish peroxidase (HRP) and biotin tyramide (biotinyl tyramid/biotin-phenol). In SPPLAT method, HRP conjugated antibody that recognizes the bait protein is added exogenously to cells. Following brief incubation with biotin-tyramide (NHS-SS-biotin-tyramide) and H_2O_2, HRP catalyzes biotin-tyramides into reactive free radicals (biotin-phenoxyl) that can attach g NHS-LC-SS biotin to the tyrosine residue of the target protein (**Figure 7**). The biotinylated proteins are then isolated by incubation of the cell lysate with streptavidin-agarose [55].

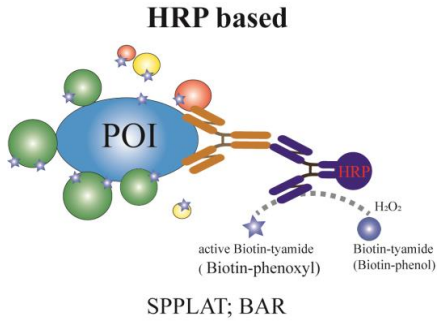


Figure 7. Schematic of HRP based proximity labeling method. A POI is recognized by the primary antibody, and HRP-coupled secondary antibody binds to the primary antibody. In the presence of H2O2, HRP catalyzes the biotin tyramide (biotin-phenol) to form reactive tyramide radicals (biotin-phenoxyl) that label proteins nearby. SPPLAT was only applied to membrane proteins. BAR extended the application to intracellular proteins in fixed cells and tissue samples.

SPPLAT has been applied to examine the proteins that co-assemble with the activated B-cell receptor (BCR) on the plasma membrane of the B-lymphocyte cell line (DT40) [54]. The results provided new insights into the composition of the B cell receptor cluster and confirmed SPPLAT as a potential tool in identifying new targets for further functional studies. A major limitation of SPPLAT is the need for suitable antibodies against bait proteins to avoid nonspecific binding.

1.3.8 Biotinylation by Antibody Recognition (BAR)

More recently, HRP based proximity labeling has been extended to fixed cells, and tissue samples named biotinylation by antibody recognition (BAR) [56] to overcome the other proximity labeling methods that cannot be easily applied on tissue samples. This method is almost identical to SPPLAT but extends the application of HRP. A primary antibody binds the bait protein in fixed cells or tissues, and a species-specific HRP-conjugated secondary antibody binds the primary antibody. After extensive washes, samples are incubated with biotin-tyramide and following brief exposure to H_2O_2 (**Figure 7**). The labeled proximal proteins and putative interactors can be subjected to MS analysis.

Researchers [56] applied this method to profile the dynamic interactome of lamin A/C in multiple fixed cell line samples and primary tissues (human, mouse) under various treatment conditions. The results indicated that the method is useful for disease pathogenesis and can be used to characterize the interactome of clinical samples from patients.

This approach does not require the generation of bait specific cell lines or animal models. Antibody-based detection prevents any artifacts related to protein fusion and overexpression. However, like SPPLAT, BAR requires a monospecific antibody that is not sensitive to fixation artifacts.

1.3.9 PUPylation-based Interaction Tagging (PUP-IT)

PUP-IT (pupylation-based interaction tagging) system was developed to identify membrane protein interactions. This system uses a small bacterial protein PafA (*Corynebacterium glutamicum* Pup ligase PafA) as the tag to fuse with POI [57]. PafA utilizes ATP to conjugate the substrate prokaryotic ubiquitin-like protein (Pup(E)) that is a small bacterial protein with Gly-Gly-Glu (GGE) motif at the C terminus, to the target protein on the lysine. In order to enrich the Pup(E) labeled proteins, Pup(E) can be fused with other protein tags, for example, biotin carboxyl carrier protein (BCCP), (BCCP-Pup(E)) that can use biotin to add biotinylation modification on GCE motif (**Figure 8**). Therefore, any protein that is modified with the GCE motif is also biotinylated and can be easily captured by the streptavidin matrix for the following purification.

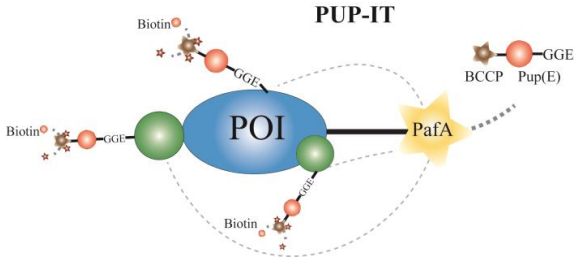


Figure 8. Schematic of PUP-IT method. The bacterial PUP ligase (PafA) is fused to the POI. PafA utilizes ATP to conjugate the Pup(E) to the prey proteins.

Unlike BioID or APEX, PafA binds the activated Pup(E) and does not allow Pup(E) to diffuse from itself freely, thus the labeling radius will be restricted to the proteins that directly interact with the bait protein. When applied a side-by-side proteomics study with BioID and PUP-IT to identify the interactor of CD28, PUP-IT provided more specific interactors than that using BioID [57]. However, PUP-IT system requires the cell to express PafA fused bait protein and the substrate protein of BCCP-Pup(E). PafA is a large enzyme (~53 kDa) and the bio-Pup(E) (~20 kDa) is also a relatively large substrate that can lead to difficulties for bait/prey proteins to go through specific membranes in cells. For this reason, PUP-IT might not be suitable to study interactions within organelles.

1.3.10 NEDDylator system

Another labeling system called the NEDDylator [58], based on the neural precursor cell expressed developmentally down-regulated protein (NEDD) 8-conjugating enzyme, Ubc12, which can covalently link NEDD8 to substrate proteins (**Figure 9A**). Since endogenous Ubc12 has very few of substrate proteins, it was fused to the engineered ubiquitin ligase, X chromosome-linked IAP (XIAP) of which RING domain has been removed. Thus, engineered XIAP will not bind the wild type of ubiquitin E2 and lose the

ability to ubiquitinate, while the fused Ubc12 can NEDDylate the substrates of XIAP (**Figure 9B**). Those NEDDylated proteins can be separated by SDS-PAGE and trypsinized for MS analysis. More than 50 potential XIAP substrates were identified and provided another powerful method to identify transient PPIs. However, the NEDDylator system requires the endogenous NEDD8 pathway (NEDD8 E1-activating enzyme, E2-conjugating enzymes, and E3 ligases) and can interfere with the actives or localization of related endogenous proteins.

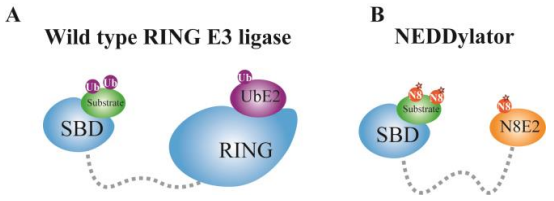


Figure 9. Wild type RING E3 ligase (A), a RING E3 ligase (blue), is bound to a ubiquitin-conjugated E2(purple), from which the ubiquitin is transferred to a lysine on the substrate (green). A substrate is attaching to the substrate domain (SBD) of substrate-binding protein. In comparison, NEDDylator (B), is engineered by removing the RING domain and replacing the ubiquitin E2 with NEDD8 E2, enabling the transfer of biotin-tagged NEDD8 (orange) to the substrate (green).

In summary, proximity-based labeling (**Table 1**) has emerged as a powerful complementary approach to classic affinity purification in the profiling of PPIs *in vivo* and *in vitro*. However, some limitations remain to be overcome by the optimization of enzyme tags to improve both temporal and spatial resolution. Additionally, some assays utilize fluorescent proteins to detect and visualize the proximal protein interaction in the live cell, such as fluorescence resonance energy transfer (FRET) (Sekar and Periasamy 2003) and bimolecular fluorescence complementation (BiFC) (Kerppola 2008), enable direct visualization of protein interactions in living cells upon spatial proximity.

Proximity labeling	BioID	BioID2	TurboID	minTurboID	2C-BioID	Split-BioID	APEX	APEX2	Split-APEX	SPPLAT	BAR	PUP-IT	NEDDylator
Enzyme	Biotin protein ligase	Biotin protein ligase	Biotin protein ligase	Biotin protein ligase	BioID and FKBP:FRB system	Biotin protein ligase	Ascorbate peroxidase	Ascorbate peroxidase	Ascorbate peroxidase	Horseradish peroxidase	Horseradish peroxidase	Pup--protein ligase	NEDD8-conjugating enzyme Ubc12
Origin	*E.Coli*	*A.Aeolicus*	*E.Coli*	*E.Coli*	*E.Coli*	*E.Coli*	*Pea* or *Soybean*	*Soybean*	*Soybean*	*Horseradish*	*Horseradish*	*C. glutamicum*	*Human*
Size (kDa)	35	27	35	28	35		~ 28	28		38	38	53	75
Mutations	R118G	R40G	Q65P, I87V, R118S, E140K, Q141R, A146Δ, S150G, L151P, V160A, T192A, K194I, M209V, M241T, S263P, I305V	N-terminal aa1-63 delated; Q65P, I87V, R118S, E140K, Q141R, A146Δ, S150G, L151P, V160A, T192A, K194I, M209V, I305V	R118G	Split BioID at aa140/141 or aa256/257	K64D, W41F, E112K	K64D, W41F, E112K, A134P	split APEX2 at aa200/201 TheN-terminal fragment (aa 1–200; K22R, R24G, G50R, K61R, H62Y, I165L, N72S, P125L, I185V); The C-terminal fragment (aa201–250)	HRP-conjugated antibody	HRP-conjugated antibody	codon-optimized for mammalian cell expression	XIAP (aa1-434) is fused via a flexible Gly-Gly-Ser-Gly linker to the NEDD8 E2, Ubc12
Substrate	Biotin and ATP	Biotin and ATP	Biotin and ATP	Biotin and ATP	Biotin and ATP; Rapamycin	Biotin and ATP	H_2O_2 and biotin-phenol	H_2O_2 and biotin-phenol	H_2O_2 and biotin-phenol	H_2O_2 and biotin-phenol	H_2O_2 and biotin-phenol	Pup(E) derivative peptides and ATP	His-biotin-tagged NEDD8
Time of labeling	16-18 h	16-18 h	10 min	10 min	16-18 h	16-18 h	1 min	1 min	1 min	5 min	10 min	24-36 h	
Labeling radius (nm)	~ 10	~ 10-15	~ 10	~ 10	~ 10	~ 10	~ 20	~ 20 (more active than APEX)	~ 25	10-200	not specific measured >10	unknown (close contact)	unknown (close contact)
Labeling residues	Lys	Lys	Lys	Lys	Lys	Lys	electron-rich amino acids such as Tyr, Trp, His and Cys	electron-rich amino acids such as Tyr, Trp, His and Cys	electron-rich amino acids such as Tyr, Trp, His and Cys	Tyr	Tyr	Lys	Lys
Application	*In vitro and in vivo*	*In vitro;in vivo(unknown)*	*In vitro and in vivo*	*In vitro and in vivo*	*In vitro;in vivo(unknown)*	*In vitro;in vivo(unknown)*	*In vitro and in vivo*	*In vitro and in vivo*	*In vitro and in vivo*	*In vitro*	*In vitro (extend the use of HRP to tissue)*	*In vitro*	*In vitro*
Screen unknown interactions	Yes	Yes	Yes	Yes	Yes	NO	Yes	Yes	No	Yes	Yes	Yes	Yes
Readout	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry; Electron microscopy	Mass sepctrometry; Electron microscopy	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry	Mass sepctrometry

Table 1: Comparison of the main features of the different proximity-tagging methods

1.4 Chemical cross-linking

Unlike AP and PL that rely on the exogenous expression of the fusion protein to characterize the interaction of the bait protein, chemical cross-linking mass spectrometry (XL-MS)[59] is an alternative technique to elucidate structural information of proteins and protein interaction networks from their native environment [60].

Cross-linking is a process that covalently links together interacting proteins by forming chemical bonds between specific residue side chains in close proximity. Crosslinking reagent (crosslinker) is typically a small organic molecule containing at least two reactive groups that are linked by spacer arm [61]. It can be hydrophobic (e.g., disuccinimidyl suberate, DSS) or hydrophilic (e.g., bis(sulfosuccinimidyl) suberate). The cross-linking step can be performed *in vivo* to determine the protein interaction profile or *in vitro* to map the topology of a protein complex [62]. After proteolytic digestion of the cross-linked sample, cross-linked residue pairs can be localized by MS-based peptide sequencing and cross-linking types (Interlink, intralink, and dead end) (**Figure 10**). However, the identification of the crosslinker labeling sites are highly challenging due to the low signal intensity of cross-linked peptides and the reliable identification of cross-linked peptides. Thus, for *in vivo* applications, most cross-linking approaches are primarily used with immune-affinity purification to determine the endogenous interaction profile of the POI [63].

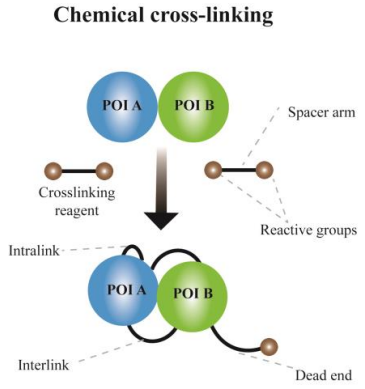


Figure 10. Simplified diagram of the chemical cross-linking. A target complex is cross-linked by the crosslinking reagent. The cross-linking types include interlink, intralink, and dead end.

XL-MS have been successfully applied to resolve the interactome of several biological systems, including viruses [64], bacteria [65], mammalian cells [66], and tissue samples [67, 68].
In summary, AP-MS is mainly used for detecting stable physical PPIs. Proximity-labeling based approaches are currently being used for mapping transient functional PPIs. XL-MS

is a powerful tool often for interpreting the 3D-structure analysis of proteins and protein complexes. All these methods serve complementary purposes for protein interaction mapping.

Besides, there are other innovative methods for the discovery of PPIs. Virotrap [69] is a “lysis free” method by fusing the POI to the HIV-1 Gag protein and trapping protein complexes in virus-like particles (VLPs) that release from mammalian cells. The VLPs can be isolated by centrifugation and filtering for later MS analysis.

2 Application of Interactome Proteomics

Correct subcellular localization of a protein is crucial for the protein function. The term spatial proteomics refers to the use of proteomics methods for subcellular protein mapping. The traditional approaches of spatial proteomics are mainly based on high throughput microscopy imaging and MS analysis of fractionated organelles. Recently developed techniques for interactome studies provide an alternative method for spatial proteomics [70]. The interactome of a single bait is a "local" spatial proteome. The combination of multiple "local" interactomes of baits from different subcellular regions can define a "global" spatial proteome of the cell.

Here, we take a few examples to review the comprehensive mapping of PPI networks to understand the spatial organization of subcellular compartments and dynamic regulation of protein complexes in signaling pathways.

2.1 Interactome of subcellular compartments

Mammalian cells are assembled by macromolecular complexes and subcellular compartment/organelles/regions. As a result, the localization of a protein provides a unique physiological environment for its specialized functions. More importantly, the interactome of a protein is changing under different physiological conditions within distinct cellular compartments. For instance, B-cell lymphoma 2 (BCL2) interacts with beclin-1 in the endoplasmic reticulum, disrupting the formation of the beclin 1- vacuolar protein sorting 34 (VPS34) autophagy-promoting complex [71]. However, BCL2 in mitochondria binds with voltage-dependent anion channel (VDAC) to regulate the release of cytochrome c during apoptosis [72]. Additionally, BCL2 can translocate into the nucleus and form the protein complex with cyclin-dependent kinase (CDK) 1, protein phosphatase 1, and nucleolin [73]. Therefore, it is crucial to understand how the spatial distribution of a protein affects its protein interactions in cell.

2.1.1 Nucleus

The nucleus is enclosed with a phospholipid rich nuclear envelope and nested with nuclear pores. Nucleoli locate inside the nucleus and are where ribosomes synthesis occurs. The nucleus works as a fundamental component to keep the genetic material separate from other activities of the cell [74].

XL-MS was used to analyze the interactome of the nucleus [75]. The intact nuclei were isolated from human U2OS cells by mechanical rupture, coupled with centrifugation. The isolated nuclei were treated with the crosslinker disuccinimidyl sulfoxide (DSSO) and followed by several sample fractionation steps at the protein/peptide level before MS analysis. Overall, 8,710 cross-links were identified, of which 2/3 represent links originating from distinct proteins (interlinks), resulting to define over 850 PPIs. These

data provided an overview of chromatin-associated PPIs and structural details of nucleoprotein assemblies.

The nuclear pore is a large complex of proteins that allows the bidirectional transport of molecules (RNA, proteins, and lips) to pass between the nucleus and cytoplasm. To understand of NPC architecture, Kim and colleagues [40] applied BirA*-tag to two conserved NPC subcomplexes (Nup107-160 complex and the Nup93 complex). They generated HEK293 cell lines that express BirA* fused bait proteins for BioID purification. The reciprocal tagging strategy and western blot confirmed the direct interaction of Nup43 with Nup85, refining the understanding of NPC organization.

In contrast to XL-MS, which requires prior purification of organelles, BioID needs one-step of protein purification to enrich the biotinylated proteins for MS analysis.

2.1.2 **Mitochondrion**

The mitochondrion is a double membrane-bound organelle consisting of inner and outer mitochondrial membranes (OMM, IMM) separated by an intermembrane space (IMS). Mitochondria play a critical role in the generation of ATP in the eukaryotic cell, and mitochondrial dysfunction can contribute to a wide range of serious diseases, including cancer [76].

Isolation of intact mitochondria is typically a laborious process, as cross-contamination from other organelles is common. Several commercial kits and manual methods [77] have been developed to isolate mitochondria from tissues or cells for proteomics studies. Among them, density gradient centrifugation has excellent advantages in obtaining good yields and functional mitochondria.

Fan and colleagues [78] have reported one of the largest surveys of mitochondrial protein interactome so far. They isolated intact mitochondria from mouse hearts using the standard density gradient centrifugation and crosslinked the mitochondrial pellets with DSSO. In total, the crosslinks uncovered a highly interconnected mitochondrial interactome with 60% of the detected crosslinks (2,041 out of 3,322) being formed between different proteins (interlinks). The obtained mitochondrial protein interactome gives insights into the architecture and sub-mitochondrial localization of defined protein assemblies.

APEX was fused to a mitochondrial matrix targeting sequence and expressed in HEK293 cells to obtain the spatial arrangement of the matrix [51]. Biotinylated proteins were subsequently enriched and identified by MS. A total of 495 proteins, including 31 not previously linked to mitochondria, were identified in the human mitochondrial matrix. Notably, six proteins previously thought to reside in the IMS or OMM were reassigned to the mitochondrial matrix by MS data, and their localizations were confirmed later by electron microscopy (EM).

2.1.3 **Primary Cilium**

Unlike nuclei and mitochondria, which are membrane-enclosed compartments, the primary ciliary membrane is contiguous with the plasma membrane, and the ciliary base is open to the cytosol.

The primary cilium is 200-300 nm in diameter and about 2-6 μm in length [79]. It plays fundamental roles in embryonic patterning or organogenesis and coordinates a series of signal transduction pathways, including Hedgehog, Wnt, Notch, Hippo, PDGFR, mTOR, and G protein-coupled receptor (GPCR) signaling pathways (Wheway, Nazlamova et al. 2018). Genetic defects in the structure or function of the human primary cilia are termed ciliopathies.

Primary cilia can be isolated by the calcium shock method or with the sucrose gradient (Raychowdhury, McLaughlin et al. 2005). Using the isolated primary cilia fraction, researchers build the proteome profile of primary cilia, including 190 proteins from mouse IMCD3 cell lines and 868 proteins from choroid plexus epithelial cells (CPECs) (Ishikawa, Thompson et al. 2012, Narita, Kozuka-Hata et al. 2012). However, the proteome data is not enough to link the ciliopathies and molecular functions of primary cilia.

The characterization of the proteins that traffic into, remove from and retain inside the cilia is the key to understand the mechanisms of ciliopathies. Several studies using proximity labeling (Mick, Rodrigues et al. 2015, Boldt, van Reeuwijk et al. 2016, Kohli, Hohne et al. 2017) have been carried out to characterize interactome of the primary cilium. APEX was fused with the ciliary targeting signal of nephrocystin-3 (NPHP3)[1–203] (cilia-APEX) and expressed in IMCD3 cell line to map the protein association in primary cilia. The result contains 162 candidate ciliary proteins without centrosomal proteins and transition zone proteins contaminations [80]. Similarly, Kohli and colleagues fused APEX2 with the cytoplasmic C-terminus of Htr6 to target APEX2 to the ciliary membrane compartment (cmAPEX) in IMCD3 cells. Their ciliary membrane-associated proteome contains 301 high confidence proteins [81]. However, the overlap of proteins between these datasets is relatively small. There are two possible explanations, i). cilia-APEX was targeted to axoneme while cmAPEX was targeted to the ciliary membrane; ii). Both studies were aiming to identify the novel protein in primary cilia and proteins were with low abundance and identified by previous studies were eluded in the final reports.

Boldt and colleagues applied Strep-FLAG on 217 ciliary proteins (91 known ciliopathy genes; 124 gold-standard ciliary proteins; 80 proteins that were frequently detected in previous proteomes studies.) expressing in HEK293 cells to probe the interactome of cilia [82]. Their landscape of primary cilia includes 1,319 proteins, 4,905 interactions, and 52 complexes. Notably, the results showed that 3M genes (*Cul7*, *Obsl1*, and *Ccdc*8) are involved in ciliogenesis, indicating that 3M syndromes could belong to ciliopathies.

Since the centrosome is at the base of the primary cilium and anchoring the cilium at the plasma membrane, many centrosome proteins are essential for ciliogenesis. Gupta and colleagues performed BioID on 58 bait proteins in HEK293 cells to map the centrosome-cilium interface (Gupta, Coyaud et al. 2015). They generated a protein topology network that contains more than 7000 interactions among 1,700 proteins. Further experiments using siRNA and high-resolution microscopy confirmed that many of these proteins are involved in centriole duplication and ciliogenesis.

2.2 Interactome of signaling pathways

Cellular signaling is the process of converting extracellular stimuli from the cell surface to the cytoplasm and the nucleus, leading to changes in the transcription of genes. The messenger transduction within the cell is dependent on consecutive protein-protein interactions. Thus, a systems-level understanding of the signaling pathway requires constructing the interaction network from PPI data. During the last decade, AP-MS has become a powerful approach for characterizing signaling pathways. Here, we review the current progress of using AP-MS to define the interaction landscape of Hippo signaling networks.

2.2.1 Hippo signaling pathway

The core Hippo pathway consists of a kinase cascade that controls cell proliferation and apoptosis. Although the interactions between key components and a few associated proteins have been studied, the knowledge of the regulation of the Hippo signaling cascade is still incomplete. In 2013, four independent interactome studies presented a comprehensive PPI network of Hippo signaling.

Kwon and colleagues generated interactome of the *Drosophila* Hippo pathway by applying AP-MS on the *Drosophila* Hippo canonical pathway components and identified 204 high-confidence interactions (HCIs) among 153 interactors. They continued to validate those interactors with functional RNAi knockdown screening and found that 102 of them affected the transcriptional activity of Yorkie [83].

Couzens and colleagues conducted a systematic study using a combination of AP-MS with FLAG-tag and BioID to reconstruct the PPI network of the human Hippo signaling pathway. They defined the network with and without inhibition of the phosphatases responsible for the dephosphorylation of the YAP/TAZ pathway by okadaic acid and ultimately generated 749 HCIs, including 599 novel interactions. The results revealed that mammalian STE20-like protein kinase (MST)- and Mps one binder kinase activator–like 1(MOB1) -family can interact in a phosphorylation-dependent manner [84].

Hauri and colleagues carried AP-MS using Strep-tag to analyze the interactome of the human Hippo pathway. The high-resolution network consists of 480 HCIs among 270 network components participating in three distinct clusters. Three major clusters include kinases, phosphatases (PP1-ASPP), and cell polarity proteins that converge at the

transcriptional coactivator, yes-associated protein 1 (YAP1). The results also discovered that YAP1 complex formation was affected by cell-cell contacts [85].

Similarly, Wang and colleagues conducted the AP-MS analysis of 32 Hippo pathway components in the HEK293 cells. They identified 550 interactions within 343 unique protein components of the Hippo pathway. The results reveal a novel interactor of YAP1, the coiled-coil domain-containing protein 85C (CCDC85C). CCDC85C binds to the WW domains of YAP and induces its nuclear export, thereby inhibiting YAP activity [86].

In conclusion, four AP-MS based studies identified many new components and modulators of the Hippo pathway. The protein interaction networks expanded the roles of the Hippo pathway in multiple biological processes and provided the great potential to discover novel biological aspects of the Hippo pathway.

3 Hedgehog signaling pathway

Hedgehog (Hh) signaling pathway was first identified and characterized in *Drosophila* [87]. Although the framework of the Hh signaling pathway is evolutionarily conserved, one of the most notable divergences between *Drosophila* and mammals is the presence of Hh ligand [88]. The *Drosophila* genome encodes a singular Hh protein, whereas there are three hedgehog homologs in mammals—Sonic hedgehog (Shh), Indian hedgehog (Ihh), and Desert hedgehog (Dhh). Among them, Shh is expressed during limb development, tissue polarity, and central nervous system development. Ihh coordinates primarily in endochondral bone development, and Dhh expression is restricted to the peripheral nervous system and gonads [89]. Because of the widespread expression of Shh, it has been most extensively studied. The fundamental knowledge of mammalian Hh signaling stems from researches on Shh.

In the absence of Hh ligand, the pathway is inactive, and its receptor transmembrane protein patched (PTC1/PTCH1) localizes within primary cilia where it actively represses the entry of smoothened (SMO). The negative regulator of the pathway, suppressor of fused (SUFU) retains the zinc-finger-containing transcription factors GLI2 and GLI3 in the cytoplasm, promoting the process of C-terminally truncated repressor forms (GLI2R and GLI3R). The processing of GLI2/GLI3 is induced by multi-site phosphorylation at the C-termini, first by protein kinase A (PKA) and subsequently by glycogen synthase kinase 3 (GSK3) and casein kinase 1 (CK1) [90] **(Figure 11A)**. Binding of Hh ligands to PTCH1 initiates the activation of the Hh pathway. PTCH1 exiting the primary cilium alleviates the inhibition of SMO and allows the translocation of SMO into the primary cilium. Thus, the Hh signal can be transmitted downstream of SMO via protein complex composed of kinesin protein (KIF7), SUFU, and GLI. KIF7 antagonizes SUFU and promotes the conversion of GLI2FL/GLI3FL into active forms (GLIA) [91]. GLIA then translocates into the nucleus and begins the transcription of the target gene GLI1[92] **(Figure 11B).**

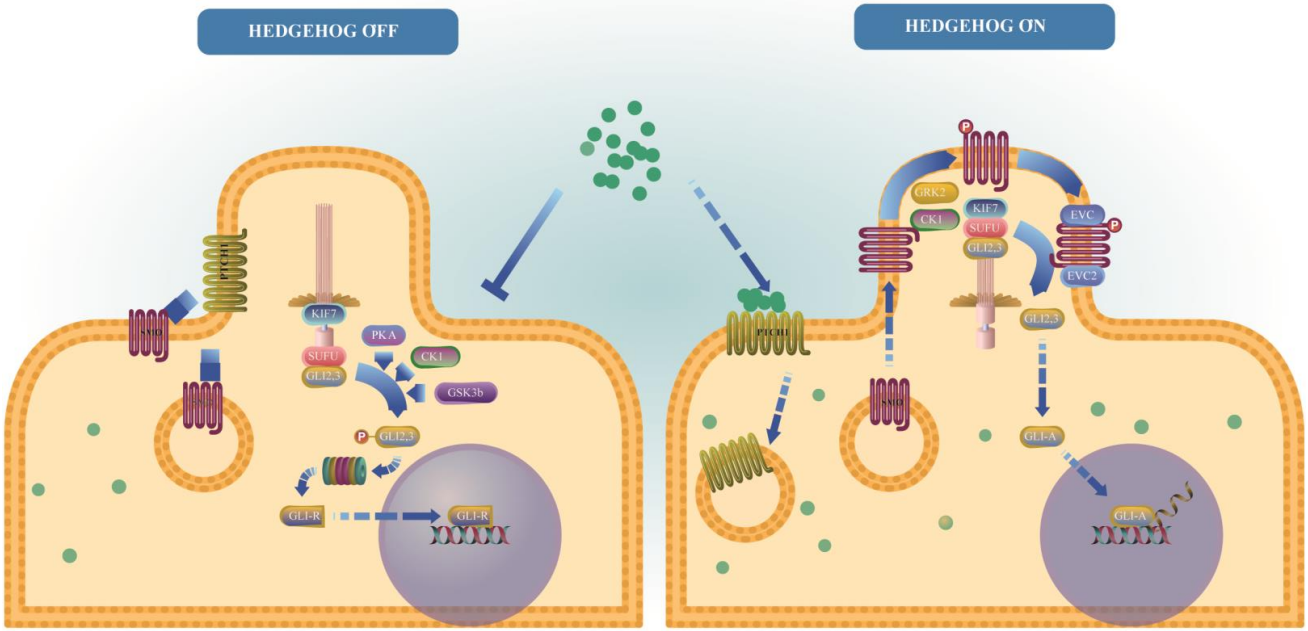

HEDGEHOG ØFF
HEDGEHOG ØN
KIF7
SUFU
GLI2,3
PKA
CK1
GSK3b
P
GLI-R
GLI-R
GRK2
CK1
KIF7
SUFU
GLI2,3
EVC
EVC2
GLI2,3
GLI-A
GLI-A

Figure 11. A simplified diagram of Hh signaling in mammals (Modified based on the Reactome pathway library [93, 94]).
(A) In the absence of the Hh ligands, the Hh receptor PTCH1 is enriched in the primary cilium to inhibit SMO signaling (via an unknown mechanism). SUFU associates with GLI2 and GLI3 to the basal body of the primary cilium, where the latter two are phosphorylated by PKA/GSK3/CK1 kinases. The phosphorylated GLI2/GLI3 will be processed into the truncated repressors GLIR with the help of SUFU/KIF7.
(B) In the presence of Hh ligands, PTCH1 exits out of cilium. As a result, SMO is phosphorylated by CK1 and the GRK2 and is enriched in the primary cilium. The activated SMO accumulates in primary cilia in association with EVC complex and prevents GLI2/GLI3 cleavage. GLI proteins dissociate from SUFU and convert into active forms (GLIA), which then translocate to the nucleus to induce target gene expression.

The Hh signaling pathway plays a crucial role in animal development and tissue homeostasis [95]. The malfunction of Hh signaling has been associated with several types of cancers (e.g., basal cell carcinoma, ovarian cancer, prostate cancer, and colon cancer)[96].

Besides the canonical Hh signaling pathway mentioned above, there are non-canonical Hh signaling pathways existing in mammals. Non-canonical Hh signaling pathways have been classified into two types earlier: Type I, PTCH1 involved signaling transduction independent of GLI activation. Type II, SMO involved signaling transduction independent of GLI activation [97-99]. Furthermore, some researchers currently consider certain crosstalk between Hh and other signaling pathways that lead to GLI transcription activation as non-canonical Hh signaling [98, 100, 101]. Mounting evidence indicates that canonical and non-canonical Hh signaling can exist together in a given cancer type. In some cases, more than one mechanism of non-canonical Hh signaling can occur [100].

Given the complexity of this signaling pathway, many mysteries are still left in the mammalian side. For example, what are the mechanisms for Hh signaling integrated with the cilia trafficking and localization? Why are cilia required for the formation of GLIR and GLIA forms? How dose PTCH regulate SMO? How dose SMO transmit the signals to the GLI transcription factors?

3.1 Cilia and Hh signaling pathway

Considering many aspects of *Drosophila* Hh signaling are conserved in mammals, it was completely unexpected when genetic experiments in mice revealed a set of proteins (IFT88, IFT172, KIF3A) were necessary for the assembly of primary cilia that were also required for Hh signaling [102, 103].

The primary cilia maintain a unique microenvironment and enable efficient molecular interactions to facilitate the Hh signal transduction. It associates Hh signaling in the following two aspects.

3.1.1 Components of Hh signaling localize in the primary cilium

Most of the proteins required for the transduction of Hh signal in mammals are enriched in primary cilia and show dynamic patterns depending on the Hh signaling state. Although the ciliary membrane is continuous with the plasma membrane, proteins within primary cilia differ significantly from those of the cell body. To date, no proof has been presented to show that proteins can be produced in the primary cilium [104]. Therefore, proteins need a transportation system to the primary cilia. The intraflagellar transport (IFT) is thought to be the main vehicle for transporting proteins into and within cilia [105]. However, there are few findings indicate that IFT transports those core components of Hh signaling into the primary cilium.

In the absence of Hh ligand, PTCH1 localizes in primary cilia [106]. Its cytoplasmic tail is necessary for ciliary localization and signaling [107]. Hh ligand binding to PTCH1 triggers its removal from the cilium, leading to the ciliary accumulation of SMO. Recently, the analysis of individual SMO molecules in live-cell using single-molecule imaging [108] confirms that SMO enters primary cilium by lateral diffusion [109] rather than active transport by IFT [110].

In the presence of Hh ligand, GLI2, and GLI3 are present at the primary cilium tip with SUFU [111]. GLI2/GLI3 and SUFU transport to primary cilia as a complex relying on IFT [112, 113]. SUFU-GLI association is proposed to be disrupted upon Hh stimulation at the cilium tip in a very short time [114]. Similar to the other core components of the Hh pathway, KIF7 also presents in the primary cilium. However, the ciliary localization of KIF7 does not require IFT [115].

3.1.2 Ciliary proteins have effects on Hh signaling

Primary cilium formation occurs during the G0/G1 phase of the cell cycle. Ciliogenesis depends on the IFT machinery and the associated Bardet-Biedl syndrome (BBS) proteins (BBSome) for the dynamic delivery of ciliary components. The IFT particles can be biochemically grouped into IFT-A and IFT-B sub-complexes [116]. The IFT-B, which comprises of 16 subunits linking to the kinesin-2 motor, is essential for anterograde transport towards the ciliary tip. In contrast, the IFT-A, which comprises six subunits plus tubby-like protein 3 (TULP3), is powered by dynein-2 motor for retrograde transport towards the ciliary basis [117]. The IFT particles carry the BBSome complex to regulate the assembly and stability of IFT particles [118]. Genetic analysis has identified many cilia-associated proteins that are required for Hh signaling transduction, and the roles of those proteins related to cilia structure and trafficking and regulating Hh signaling are varied.

Mutations that prevent ciliogenesis can cause abnormal Hh signaling, such as those disrupting the IFT-B protein complex (IFT52, IFT57, IFT88, and IFT172). Mutations of IFT88 and IFT172 in mice, lead to loss of several ventral cell types and reduced expression of PTCH1 [103]. Mutations of IFT88 and IFT172 in humans that can cause

retinal degeneration showed mislocalization of GLI1 compared with the controls [119, 120]. Similarly, mutations in IFT52 and IFT57 that show the loss of motor neuron specification in the ventral neural tube could disrupt Hh signaling [121, 122].

In contrast to the mutations leading to completed loss of primary cilia, there are mutations that partial defects in primary cilia structure to affect Hh signalings, such as dynein cytoplasmic 2 heavy chain 1(DYNC2H1) and ADP ribosylation factor Like GTPase 13B (ARL13B). DYNC2H1 mutations are the common cause of Jeune syndrome in humans [123]. The DYNC2H1 mutant MEF cells have short primary cilia and abnormal bulges at the tip of the axoneme [124]. Although SMO is constitutively present in cilia in these cells, the presence of SMO is not sufficient to activate the Hh signaling, and the ciliary localization of SMO could not be further increased by the addition of SAG (SMO agonist, CAS 912545-86-9) [125]. Mutations of ARL13B cause Joubert syndrome in humans too [126]. In mouse MEF cells, a null mutation of ARL13B leads to short cilia and reduces Hh signal transduction [127].

Lastly, there are mutations of ciliary proteins that do not affect ciliogenesis but still affect Hh signaling, such as TULP3 and limbin (EVC2). TULP3 binds to the IFT-A complex to promote ciliary GPCRs trafficking. TULP3 regulates polycystic kidney disease in a cellular context-dependent manner [128]. It localizes to the ciliary base, and knockdown of TULP3 does not lead to any ciliation defect [117]. However, mice lacking the TULP3 developed abnormalities of both the neural tube and limbs consistent with aberrant Hh signaling [129]. Ciliary transmembrane proteins, EVC (Ellis-van Creveld syndrome protein) and EVC2 are responsible for the human skeletal disorder Ellis-van Creveld syndrome. Upon pathway activation, EVC-EVC2 complex binds with SMO at EVC zone within cilium [130]. Cells lacking EVC/EVC2 appear structurally normal primary cilia [131]. In EVC2-deficient chondrocytes, SMO translocation to the cilium was typical in the presence of SAG, but with reduced protein levels of PTCH1, GLI1, and GLI2 [130].

3.2 Interactome of Hh signaling pathway

The core components of Hh signaling in mammals include Hh ligands (Hh), the receptor PTCH, signal transducer SMO, regulator SUFU, KIF7, and the transcription factors GLIs (GLI1, GLI2, GLI3). As an interesting curiosity, the specific characteristic of the Hh pathway is a large number of subsequent interactions with upstream components negatively regulating the downstream pathway components. Here, we focus on recent updates on a series of PPIs around the core components of this pathway to understanding Hh signaling in development and cancer.

3.2.1 Protein-protein interactions of Hh ligand

The Hh gene encodes a precursor protein that undergoes the auto-cleavage process to generate an amino-terminal signaling peptide (HhN). During this process, the C-terminus of HhN is binding to a cholesterol moiety, and the N-terminal cysteine residue is modified

by palmitoylation in the endoplasmic reticulum. The three Hh ligands Shh, Ihh, and Dhh, are similarly processed, modified, and released [132].

When Hh ligand reaches and attaches to the plasma membrane, transmembrane protein dispatched (DISP), and scube 2 (SCUBE2) can interact with the cholesterol anchor of Hh ligand through different structural aspects. These interactions keep the cholesterol anchor of Hh from the aqueous environment and thus keep the Hh ligand soluble and to be released from the producing cells [133]. Once Hh monomers are released from the producing cells, they can form the multimeric complexes by interacting with the heparan sulfate proteoglycans (HSPGs), which can pass to more distantly located cells. HSPGs may positively or negatively affect Hh ligand in presence on the cell surface or tissue distribution [134, 135].

Although the major receptor against Hh is PTCH1, there are additional membrane proteins, hedgehog interacting proteins (HHIP), growth arrest-specific 1 (GAS1), CAM-related/down-regulated by oncogenes (CDO) and brother of CDO (BOC) that bind Hh.

HHIP is a membrane glycoprotein acting as competitors of PTCH1 to bind Hh ligands with an affinity equivalent to the affinity between Hh and PTCH1 [136]. Thus, the Hh ligand is unavailable to PTCH1, and the Hh signaling pathway is negatively regulated by HHIP.

Unlike HHIP function as an antagonist of the Hh pathway, BOC, CDO, and GAS1 act positively on Hh signaling. BOC and CDO are single-pass transmembrane proteins of the immunoglobulin superfamily, which are conserved from *drosophila* to mouse. BOC and CDO contain multiple immunoglobulins and fibronectin type III (FNIII) repeats. BOC and CDO both bind Hh ligand with their FNIII repeats, and these interactions are necessary for the enhancement of Hh signaling [137]. Moreover, biochemical and crystallographic studies show that interactions between the BOC/CDO with Hh ligand are calcium-dependent and sensitive to relevant pH changes [138]. GAS1 is a glycosylphosphatidylinositol (GPI)-linked membrane protein in mammals [139]. Unlike BOC and CDO, there is no orthologues of GAS1 in *Drosophila* [140]. N-glycosylation of GAS1 is required for the interaction between GAS1 and Hh ligand [141]. GAS1 positively regulates Hh signaling in multiple developmental contexts, especially at regions where Hh works at low concentration [142]. More than 1% of patients with holoprosencephaly show sequence variants in GAS1 that can impair its physical interaction with Hh ligand, suggesting the necessity of GAS1 for Hh-dependent forebrain patterning. BOC, CDO, and GAS1 play overlapping and essential roles during Hh-mediated mammalian ventral neural patterning. The mice lacking CDO, BOC, and GAS1 individually show no effect or relatively mild effects on Hh-dependent neural patterning. However, in mice lacking any two of these proteins function together, Hh-dependent motor neural progenitor specification is severely impaired (Allen et al., 2011; Izzi et al., 2011).

There are a growing number of cell surface receptors that can bind the Hh ligand, highlighting the complex and tightly regulated nature of Hh signaling. Future work is required to assess whether all these receptors functionally cooperate or compete with each other for Hh binding at the cell surface.

3.2.2 Protein-protein interactions of PTCH

PTCH contains 12 hydrophobic transmembrane segments (TMs), two large hydrophilic extracellular domains (ECDs), and a C-terminal domain (CTD) [143]. The TMs 2-6 constitute the sterol-sensing domain (SSD) that is a phylogenetically conserved domain existing in many of sterol transport and metabolism-related proteins. CTD contains the "PPXY" motif (endocytic sorting signals) to direct PTCH to a degradative pathway [144]. Furthermore, PTCH lacking the PPXY motif displays altered intracellular localization [145] **(Figure 12).** In mammals, there are two homologous PTCH receptors, PTCH1 and PTCH2, both of which can bind all the Hh ligands (Shh, Ihh and Dhh) with similar affinity and share overlapping functions in limb development [143, 146]. Additionally, co-localization and heteromeric interaction between the PTCH1 and PTCH2 have been detected in double-transfected cells [147].

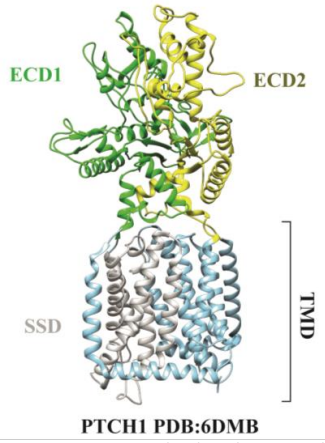


Figure 12. Cryo-EM structures of human PTCH1[148]. The domain composition is highlighted with corresponding color showing in structure.

However, PTCH1 and PTCH2 have different functions based on their expression during epidermal development [149]. PTCH1 is expressed throughout the mouse embryo, and PTCH2 is preferentially expressed in the skin and testis [147]. Because of the essential role of PTCH1 in development, most published studies focus on PTCH1, while the role of PTCH2 remains ambiguous.

In the absence of Hh ligand, PTCH1 resides in punctate structures along the ciliary membrane [150]. PTCH2 also localizes within the ciliary membrane. However, this localization is not dependent on the interaction with PTCH1 [151]. Hh ligand binds to the PTCH1 surface constituted by the upper helical domains of the two ECDs in a 2:1 ratio. The binding interfaces between one Hh ligand and two PTCH1 are asymmetric [148, 152, 153].

In the presence of Hh ligand, both PTCH1 and PTCH2 can form the distinct complexes with Hh co-receptors BOC, CDO, and GAS1 [151, 154, 155]. Hh ligand binding to PTCH1 causes the internalization of the Hh-PTCH1 complex through endosomes. Endosomes sort PTCH1 to be recycled back to the plasma membrane or to late endosomes and lysosomes for degradation [156]. This process is initiated by PTCH1 ubiquitination. Smad ubiquitin regulatory factor 1 (SMURF1) and SMURF2 are E3 ubiquitin ligases and share redundant functions during development. Both can bind the PPXY motif of PTCH1 to mediate PTCH1 ubiquitination upon binding to Hh [157]. However, basal PTCH1 turnover in the absence of Hh ligand is mediated by ubiquitin E3 ligases Itch and WWP2. PTCH1 is ubiquitylated at K1413 in the C-terminal domain by Itch and, to a lesser extent, by WWP2, to target for proteasomal degradation [158]. Hh ligand binding to PTCH1 relieves the inhibition of SMO to activate the signal transduction. However, PTCH1 does not physically interact with SMO [159], and the mechanism by which PTCH represses SMO is not entirely understood. It has been suggested that PTCH inhibits SMO in a catalytic manner by pumping a sterol-like regulator, considering its SSD sharing homology with a class of bacterial transporter [160, 161].

Despite PTCH working as the principal receptor of Hh ligand, it also modulates other non-canonical Hh signaling cascades such as apoptosis and cell proliferation [99].

In the absence of Hh ligands, PTCH1 induces apoptosis through its conserved Asp residue (D1392) in the CTD [143, 162] which serves as a caspase cleavage site to expose the binding site for a protein complex. This protein complex contains down-regulated in rhabdomyosarcoma LIM domain protein (DRAL), tumor up-regulated CARD-containing antagonist of caspase-9 (TUCAN) and the E3 ligase NEDD4 that activates caspase 9 via its ubiquitination [163, 164]. Moreover, NEDD4 constitutively interacts with the CTD of PTCH1 independent of PPXY motifs [164] and mediated potentiation of PTCH1-induced caspase-3 activation, leading to an irreversible commitment to cell death [165]. The immunoglobulin superfamily adhesion molecule close homolog of L1 (CHL1) can prevent PTCH1 induced cell death by binding the first extracellular loop of PTCH1in an SMO-dependent mechanism [166].

In the absence of Hh ligands, PTCH1 interacts with constitutively phosphorylated cyclin B1 (CCNB1) derivatives in the cytosol, which is essential for the control of the cell cycle [167]. The addition of the Hh ligand induces a conformation change of PTCH1 and increases the binding between G protein-coupled receptor kinase 2 (GRK2) and PTCH1,

leading to the displacement of cyclin B1. Thus, cyclin B1 can translocate to the nucleus and complete the mitosis [168].

Hence, concerning the essential roles of PTCH, further investigations are needed to understand whether there are intermediate chemicals/ proteins between PTCH and SMO, and how several cellular and molecular processes drive and regulate the Hh signaling.

3.2.3 **Protein-protein interactions of SMO**

SMO belongs to the Class Frizzled of the GPCR superfamily. It has a long extracellular N-terminal domain containing a conserved cysteine-rich domain (CRD), three extracellular and three intracellular loops (ECL and ICL), followed by a small linker domain (LD), which connects to the seven transmembrane domains (TM) and an intracellular CTD [169] (**Figure 13**).

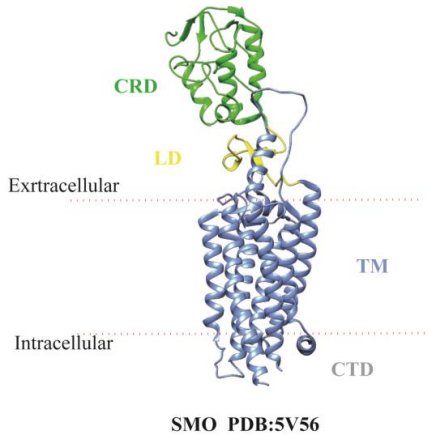


Figure 13. Cryo-EM structures of human SMO [170]. The domain composition is highlighted with corresponding color showing in structure.

SMO exists as constitutive dimers/oligomers. In the absence of HH, CTD of SMO assumes a "closed" inactive conformation to prevent the SMO association, whereas its extracellular N-terminus forms a constitutive dimer [171]. In the presence of Hh ligand, SMO interacts with disks large homolog 5 (DLG5) by its CTD at the ciliary base and translocates into the primary cilium. However, SMO trafficking into the primary cilium is independent of its interaction with DLG5 [172]. Its ciliary localization is regulated by phosphorylation. Hh ligand presence increases casein kinase 1 (CK1) and GRK2 binding with SMO and leads to a hyperphosphorylated CTD form of SMO. These phosphorylation events prevent the ubiquitination of SMO and induce a dimerization/oligomerization of SMO in the CTD region [173]. The process of de-ubiquitination of SMO is controlled by binding with ubiquitin-specific protease 8 (USP8)

[174]. Several additional proteins, KIF3A, β-arrestin (ARRB) [175], and integrin-linked kinase (ILK) [176], can interact with the phosphorylated SMO to contribute its ciliary accumulation. Within the primary cilia, SMO interacts with the EVC-EVC2 in the EvC Zone [177, 178]. The IQ-domain containing protein E (IQCE) and EF-hand calcium-binding domain-containing protein 7 (EFCAB7) can anchor the EVC-EVC2/SMO complex at near base of the cilia (EvC zone) [179]. Contrary to previous believes that signal transduction required SMO at the tips of the cilium [180], SMO is tethering at the EvC zone to transduce the Hh signal. However, further work is required to find how these interactors can collaborate to regulate SMO ciliary localization and signal transmission.

SMO activation also can involve in the non-canonical Hh signaling, which is unrelated to GLI-mediated transcription, but the mechanisms are not yet clear. SMO, as a member of the frizzled (class F) family of GPCRs, shows the interaction with specific heterotrimeric guanine nucleotide-binding protein (G proteins) [181]. Presence of Hh ligand can increase the recruitment of Gαi to SMO and increase the GTPase activity of Gαi, to modulate Ca2+ flux, RhoA and Rac activation [182-185]. Interestingly, SMO association with Gαi could contribute to the survival of diffuse large B-cell lymphoma cells (DLBCL) and subsequently activate nuclear factor NF-κB [186].

Several synthetic and naturally occurring small-molecules that interact with SMO, and function as agonist /antagonist are extensively reviewed elsewhere [161, 187, 188]. To date, the US food and drug administration (FDA) has approved two inhibitors of SMO, vismodegib, and sonidegib, for the treatment of basal cell carcinoma (Li, Song et al. 2019). Recent researches also suggested that PTCH1 might regulate SMO localization and activation through oxysterol or cholesterol [161]. They both can work as positive regulators of SMO in the Hh pathway downstream of PTCH1 [161, 189, 190]. However, it remains to be determined whether oxysterol and cholesterol act as second messengers in Hh signaling.

3.2.4 **Protein-protein interactions of SUFU**

SUFU functions in the downstream of SMO and is an essential negative regulator of Hh signaling by modulating the transcriptional activity of GLI. Mutations in human SUFU are associated with multiple cancer forms, including medulloblastoma and basal cell carcinoma syndrome [191]. SUFU consists of an N-terminal domain (NTD), a middle intrinsically disordered region (IDR), and a CTD **(Figure 14)** [192].

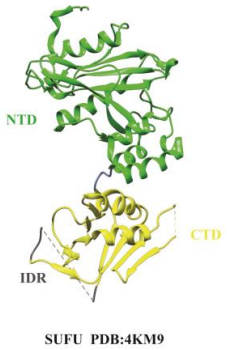


Figure 14. Cryo-EM structures of human SUFU [191]. The domain composition is highlighted with corresponding color showing in structure.

Without Hh ligand, Both SUFU-NTD and SUFU-CTD are involved in binding with GLIs. Hh treatment induces the dissociation of SUFU from GLIs, the conformation of SUFU becomes in an 'open' state in which NTD and CTD are further away [191, 192].

When Hh signaling is "off" in the cytoplasm, SUFU interacts with cAMP-dependent protein kinase A (PKA) and glycogen synthase kinase-3 beta (GSK3β) that can sequentially phosphorylate SUFU at Ser-346 and Ser-342 [113]. NIMA-related expressed kinase 2A (NEK2A) can also interact with SUFU to prevent it from proteasome-dependent degradation [193]. The phosphorylated SUFU is stable in the cytoplasm/cilia and sequesters GLI2/3 proteins in the cytoplasm/cilia to regulate their transcriptional activity [113, 150].

When Hh signaling is "off" in the nucleus, SUFU, GLIs, transcriptional repressor p66- β (P66β), and a Myc-binding protein (MYCBP) form a protein complex to inhibit GLI-mediated Hh target gene expression. However, P66β negatively and MYCBP positively regulate Hh signaling. MYCBP can remain inactive in the absence of Hh ligand [194]. SUFU can also associate with Sin3-associated polypeptide 18 (SAP18), Galectin3, and other nuclear proteins to form a ternary DNA-binding complex on a GLI DNA-binding site to repress transcription [195, 196].

When Hh signaling is "off" in primary cilia, SUFU promotes the formation of the GLI3R. The ubiquitylation of SUFU mediates this process. The HECT E3 ubiquitin ligase Itch, in complex with the adaptor protein ARRB2, binds SUFU and leads SUFU K63-linked ubiquitylation to increase the interaction between SUFU and GLI3. Consequently, the amount of GLI3R is keeping at a certain level to turn the Hh pathway off [197].

When Hh signaling is "on" in the primary cilium, the phosphorylated SUFU forms the complex with GLI2/3 traveling to the tip of the primary cilium. The dephosphorylation

of SUFU triggers its retrograde export. It is not yet clear which phosphatase is responsible for the dephosphorylation of SUFU in mammals [113].

When Hh signaling is "on" in the nucleus, the dephosphorylated SUFU initiates the interaction with E3 ubiquitin ligase F-box and leucine-rich repeat protein 17 (FBXl17) to ubiquitinate its lysine 257. The ubiquitylated SUFU allows the release of GLI1 from SUFU [198]. Pathway activation promotes SUFU/p66β dissociation from GLIs (possible GLI1/2) and enhances MYCBP-GLI interaction to promote GLI-mediated gene expression [194].

In addition to the interaction between GLIs, SUFU can interact with GLI-similar (GLIS) proteins that constitute a subfamily of Krüppel-like zinc finger proteins [199]. SUFU binds the VYGHF motif within the N-terminal conserved region of GLIS3. SUFU-GLIS3 interaction inhibits GLIS3 ubiquitination and enhances GLIS3 stability to inhibit GLIS3-mediated activation of the insulin promoter [200]. In contrast to GLIS3, SUFU, and GLIS2 interaction does not appear to influence GLIS2 stability or transactivation function [201].

3.2.5 **Protein-protein interactions of KIF7**

KIF7 is a member of the kinesin-4 superfamily. Many kinesins function as homodimers [202]. As a typical kinesin, homodimeric KIF7 possesses an N-terminal globular motor/head domain that contains nucleotide-binding and microtubule interacting regions, followed by a neck/stalk domain predicted to form a discontinuous coiled-coil for flexibility of the motor domain and a globular C-terminal tail for binding of cargos [203, 204] (**Figure 15**). The motor domain and the first coiled-coil of KIF7 serve as the sites of homodimerization [205].

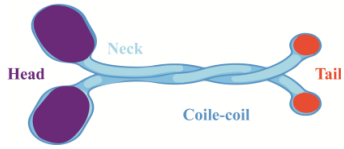


Figure 15. Schematic representation of KIF7. The domain composition of the KIF7 is highlighted (Modified from [206]).

KIF7, a vertebrate homolog of Cos2, was considered only crucial for Hh signaling in zebrafish [207]. Later, using KIF7 knockout mice, researchers showed that KIF7 played an essential role in mammalian Hh signaling [208]. KIF7 also acts downstream of SMO at a level similar to Sufu but may function separately from SUFU [209].

In the absence of Hh ligand, KIF7 is enriched at the ciliary base in MDCK, MEF, IMCD3, and 3T3 cell lines. It moves from the base to the ciliary tip in response to the stimulation of Hh [172, 205, 210]. Another research demonstrated that KIF7 was associated with one

of the two centrioles before ciliogenesis and localized to the tips of primary cilia at all stages of ciliogenesis in MEFs [115]. Thus, it can modulate Hh signaling at least partially by inducing tubulin depolymerization. Discrepancies between the localizations obtained from different methods and cell lines represent a complex distribution and function of the KIF7 within primary cilium. Indeed, KIF7 could play negatively and positively regulatory roles in Hh signaling in different tissues. KIF7 promotes Hh pathway activity in chondrocytes during growth plate development, whereas it inhibits Hh signaling during early embryogenesis [208, 211].

Co-immunoprecipitation experiments using transfected lysates revealed that GLI2/3 could bind to the motor domain of KIF7[212], suggesting KIF7 may act as the cofactor for the ciliary localization of GLI-SUFU. However, in *Kif7*$^{-/-}$ cells, GLI2/3 and SUFU can still accumulate in cilia in response to Hh ligands, indicating that KIF7 likely cannot be the motor that transports GLIs from the base to the tip of cilia [115]. Upon pathway activation, SMO forms a protein complex with EVC-EVC2 in cilia. This protein complex requires the recruitment of KIF7 [177] and promotes the interaction between the PPF1A-PP2A complex and KIF7 [206]. Thus, the dephosphorylated KIF7 travels together with GLIs to the tips of cilia, leading to SUFU-GLIs dissociation. The scaffolding protein DLG5 has been reported to interact with KIF7 at the base of the primary cilium to facilitate the accumulation of KIF7 at ciliary tips [172].

The mechanism by which KIF7 controls cilia length [115], and in turn, contributes to pathway activation is not clear. Whether KIF7 is required for GLIs-SUFU trafficking and disassociation in cilia during pathway activation remains to be determined.

3.2.6 **Protein-protein interactions of GLIs**

In *Drosophila*, cubitus interruptus (Ci), acts as both an activator and repressor to regulate gene expression of Hh signaling. In mammals, there are three homologs of Ci: GLI1, GLI2, and GLI3. GLIs belong to the family of Kruppel-like transcription factors. They contain a highly conserved C2H2-Kruppel-type zinc-finger motifs in their DNA binding domains **(Figure 16)**. However, each of GLIs exhibits distinct transcriptional activity from others.

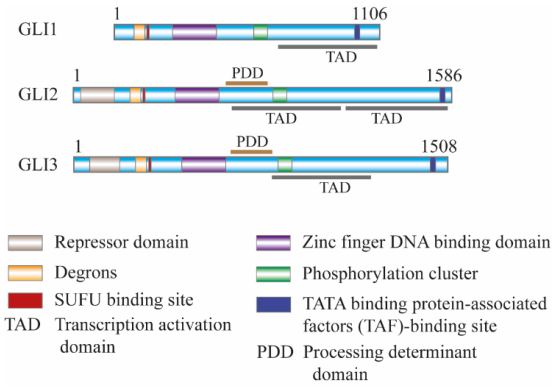


Figure 16. Schematic representation of GLI isoforms. The domain composition of the three GLIs proteins determines their function (modified from [213]).

Both full-length GLI2 and GLI3 are transcription activators (GLI2A/GLI3A). The N-terminals of GLI2/GLI3 contains the repressor domain. The processing determinant domain (PDD) determines GLI2/GLI3 to be proteolytically truncated the C-terminal trans-activating domains (TADs) (Figure 6), to form transcription repressor (GLI2R/GLI3R). The formation of both GLIA and GLIR are dependent on primary cilia [121, 214]. GLI2 and GLI3 cooperate to modulate the target gene expression. GLI1 works as part of a positive feedback loop to amplify Hh signaling due to the lack of the N-terminal suppressor domain (Figure 6). However, GLI1 is not required for Hh signaling transduction in mammals [215], and its expression is a readout of Hh signaling [216].

In the absence of Hh, activities of GLIs are restrained by phosphorylation. Phosphorylation GLI2/3 by PKA at phosphorylation cluster (six conserved serine residues) facilitates them for subsequent phosphorylation by casein kinase 1 (CK1) and GSK3β [217]. The hyperphosphorylated GLI2/3 protein creates binding sites for SCF (SKP1, Cullin, F-box containing complex) E3 ligase (βTrCP) [218], which in turn ubiquitinates GLI2/3 for partial proteolytic cleavation to generate the GLI2R/GLI3R. The degradation of activated GLI2/3 is controlled by another E3 complex cullin3- Speckle-type POZ protein (SPOP) [145]. However, GLI1 is not a substrate of cullin3-SPOP, and its degradation is mediated by Numb and E3 ligase Itch [219].

Chromatin-associated SAFB-like transcription modulator (SLTM) can interact with all three GLIs to regulate Hh signaling [220]. It implies that the relative expression levels of GLIA and GLIR determine whether SLTM works as an activator or a repressor of Hh signaling. In the absence of Hh ligand, SLTM in the nucleus facilitates the binding of GLI3R to GLI1 regulatory region and enhances the repressor function of GLI3R. While the presence of Hh ligand, SLTM cooperates with GLIA to mediated target gene activation and cell differentiation [221].

In addition to the activation of GLI by the canonical Hh pathway, non-canonical mechanisms contribute to regulating the transcriptional activation of GLI genes as well.

For instance, the beta subunit of the IκB kinase (IKK) complex (IKKβ) can increase GLI1 protein levels and transcriptional activity in DLBCL by forming the protein complex with GLI1 to mediate its phosphorylation level [222]. Furthermore, GLI1 interacts with SMAD proteins directly to modulate transforming growth factor β (TGFβ) pathway target gene expression in a p300/CREB-binding protein-associated factor (PCAF)-dependent manner [223].

Recent discoveries of protein interactions that involve in the regulation of Hh signaling have uncovered new mechanistic and biological insights in Hh signal transduction. However, these PPIs have mainly been measured in specific tissues or cell lines. The data from one research may differ from another result. These discrepancies are due to the context of tissues, cell lines, disease conditions, and even the methodological approaches they applied, making it difficult to interpret these interactions at a universal level. Therefore, a complete understanding of the mammalian Hh signaling interactome will require the experimental definition of the context-specific interaction networks across cell types and tissues, and their dynamic changes in response to Hh stimulation, differentiation, and disease.

II AIMS OF THE STUDY

The primary goal of this thesis was to set up a workflow that can be used to comprehensively study PPIs in different compartments and conditions, with the special interests in understanding the PPIs in mitochondria and Hh signaling pathway.

The individual aims of the presented studies were as follows:

- Utilizing both AP-MS and BioID to generate a versatile workflow (MAC-tag system) that could be used for the identification of stable physical and transient functional PPIs. (I-III manuscript)
- MAC-tag application on of sub-organelles (mitochondria) (I-II manuscript).
- MAC-tag application on the characterization of the signaling pathway (Hh signaling pathway) (IV manuscript)

III MATERIALS AND METHODS

1 DNA constructs

The Gateway® recombination technology (Thermo Fisher Scientific) was utilized for creating all the DNA constructs. The entry vectors of all studied genes were either gathered from human ORFeome collection or gene synthesis service (Details were described in related publications). They were cloned into either N- or C terminal of the following destination vectors (**Table 2**).

Destination Vector	Resource	Study
MAC-tag-N	Addgene Plasmid #108078	I,II,IV
MAC-tag-C	Addgene Plasmid #108077	I,II,III,IV
pcDNA5/FRT/TO/StrepIII/HA/GW-N	Varjosalo lab	I,IV
pcDNA5/FRT/TO/StrepIII/HA/GW-C	Varjosalo lab	I,IV
pcDNA5/FRT/TO/BirA/Myc/GW-N	Varjosalo lab	I,IV
pcDNA5/FRT/TO/BirA/Myc/GW-C	Varjosalo lab	I,IV
pcDNA3.1 MCS-BirA(R118G)-HA	Addgene Plasmid #36047	II

Table 2: List of destination vectors used in the presented studies. (FRT: Flp-FRT recombination technology; TO: TetO2 promoter).

2 Cell Lines and Cell Culture (I-IV)

The experiments were conducted in the following cell lines (**Table 3**):

Cell line	Catalog number	Culture condition	Organism	Study
Flp-In™ 293 T-REx cell line	R78007 (ThermoFisher)	DMEM supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Homo sapiens	I,III,IV
Cervical cancer cell line Hela	ATCC® CCL-2™	DMEM supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Homo sapiens	I
Osteosarcoma cell line U-2 OS	ATCC® HTB-96™	RPMI 1640 supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Homo sapiens	I
Prostate cancer cell line DU145	ATCC® HTB-81™	McCoy's 5A Medium supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Homo sapiens	I

Bone osteosarcoma (143B) cell line	ATCC® CRL-8303™	DMEM supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Homo sapiens	II
Flp-In™-3T3 cell line	R76107 (ThermoFisher)	DMEM supplemented with 10% FBS,2 mM L-glutamine 1% Pen-Strep	Mus musculus	IV
Flp-In™ MDCK T-REx cell line	A gift from Prof. Pekka Lappalainen	MEM supplemented with 10% newborn calf serum, 2 mM L-glutamine 1% Pen-Strep	Canis familiaris	IV

Table 3: List of cell lines and culture media used in described studies. (DMEM= Dulbecco's modified Eagle medium; MEM= minimum essential medium; FBS: Fetal bovine serum; Pen-Strep: Penicillin Streptomycin).

3 Liquid Chromatography (LC) and Mass spectrometry (MS)

The peptide mixtures were separated via LC, ionized by electrospray, and analyzed by mass spectrometry (**Table 4**).

Liquid chromatography	**Mass spectrometry**	**C18 pre-column**	**C18 analytical column**	**Study**
nano LCII Liquid Chromatograph	Orbitrap™ Elite Velos Pro™	2 cm x 100 µm, 5 µm,120 Å, 500 bar	10 cm x 75 µm, 3 µm,120 Å, 500 bar	IV
EASY-nLC 1000 Liquid Chromatograph	Orbitrap™ Elite Velos Pro™	2 cm x 100 µm, 5 µm,100 Å, 1200 bar	15 cm x 75 µm, 3 µm,100 Å, 1200 bar	IV
EASY-nLC 1000 Liquid Chromatograph	Q Exactive™ Hybrid Quadrupole-Orbitrap™	2 cm x 100 µm, 5 µm,100 Å, 1200 bar	15 cm x 75 µm, 3 µm,100 Å, 1200 bar	I, II, III, IV

Table 4: List of mass spectrometers used for analysis

The solvent gradient increased linearly from 5%-35% acetonitrile in 0.1% formic acid at a flow rate of 300 nL/min for either 60 min (I, II, III, IV) or 90 min (IV).

4 Data Mining and processing

Various resources and software (**Table 5**) have been used for data analysis, interpretation, presentation, and visualization.

Resource/ software	Function	Link	Study
Cytoscape	Open source software platform for visualizing complex networks	https://cytoscape.org/	I, II, IV
PINA2.0	Database for PPIs	http://omics.bjcancer.org/pina/	I, II, IV
Intact	Database for PPIs	https://www.ebi.ac.uk/intact/	I, IV
DAVID	Gene ontology analysis	https://david.ncifcrf.gov/	I, II, IV
CORUM	Database for mammalian protein complexes annotation	https://mips.helmholtz-muenchen.de/corum/	I, IV
Maxquant (1.6.2)	Identify and quantify proteins in complex biological samples for analyzing large-scale mass-spectrometric data sets	https://www.biochem.mpg.de/5111795/maxquant	I, IV
Proteome Discoverer™ Software (V1.4.0)	Identify and quantify proteins in complex biological samples and contains database search algorithms (SEQUEST)(commercial software)	https://www.thermofisher.com/order/catalog/product/OPTON-30795	I, II, III, IV
Xcalibur™ Software	Data acquisition on Thermo Scientific™ LC-MS systems and related instruments (commercial software)	https://www.thermofisher.com/order/catalog/product/OPTON-30487	I, II, III, IV
SAINTexpress	Significance Analysis of INTeractome (SAINT) is the software used for assigning confidence scores to PPIs	http://saint-apms.sourceforge.net/Main.html	I, III, IV
Tree View	open-source Java app for visualizing large data matrices	http://jtreeview.sourceforge.net/	I, II, IV
matrix2png	Heatmap generation	http://www.chibi.ubc.ca/matrix2png/	I, II, IV
ProHits-viz	data visulization	https://prohits-viz.lunenfeld.ca/	IV
R	R is a software for statistics and graphics.	https://www.r-project.org/	I, II, III, IV
Python	Python is a programming language for processing large data set	https://www.python.org/	I, II, III, IV
PeptideAtlas	Mass spectrometry data deposition	http://www.peptideatlas.org/	I,IV
Adobe Photoshop CC	Data presentation andn visulization (commercial software)	https://www.adobe.com/	I, II, III, IV
Image J	Software package for image processing	https://imagej.net/	I, III, IV
Leica Application Suite X	Software on Leica Microsystems confocal systems for image obtaining and processing	https://www.leica-microsystems.com/	I, III, IV

Table 5. List of databases and software used for the studies.

5 Other assays and methods

Other assays and methods that have been used can be found from the original publications (I-IV), as indicated in Table 6 below.

Method	Publications
Statistics	I, II, III, IV
Immunofluorescence	I, III, IV
Western Blotting	I, III
Luciferase Report Assay	IV

Table 6. List of assays and methods used in the thesis.

IV RESULTS AND DISCUSSION

1 The workflow of the MAC- tag approach

AP-MS and BioID have made significant contributions to proteomics studies. Recently, several studies revealed their complementary nature [206, 224]. However, PPIs analysis using AP-MS and BioID is still time- consuming and resource-intense in respect to cell line generation and protein complex purification. To overcome these caveats, we presented here a Gateway®-compatible MAC (Multiple Approaches Combined) -tag containing the Strep-tag for affinity purification and BirA* for proximity labeling into a single construct (**I, Supplementary Figure 1 a)**. Subsequently, a single pool of *isogenic*, *inducible* cells of POI can be rapidly established with the help of the Flp-In™ T-REx™ system [225]. Strep-Tactin® matrix enables one-step protein purification [226] in the following chromatography step that can be parallel applied on both AP-MS and BioID pipelines (**I, Figure 1; III, Figure 1**). The two pipelines differ in cell culture and cell lysis stages. For the BioID pipeline, the activation of the BirA* requires the addition of exogenous biotin to the culture media (**I, Supplementary Figure 1b,c**), and the harsh solubilizing condition is applied to release all the biotinylated proteins (**III, Figure 3**). After steps of purification, the specific biotinylated proteins are enriched in the BioID samples (**III, Figure 2 d**).

The integrated workflow significantly enhances the throughput of the stable cell line generation, facilitates the parallel implementation of both AP-MS and BioID with high sensitivity and reproducibility (**I, Supplementary Figure 1 d,e**).

1.1 Validation of the correct localization of the MAC-tagged bait proteins and their biotinylated interactors

Since the subcellular localization of a protein is tied to its function, it is crucial to show the MAC-tagged protein in the correct subcellular localization. We used fluorescence microscopy to assess the localization of 18 MAC-tagged cellular localization markers (**I, Supplementary Data 1 a**). The tagged-localization markers can be detected by the anti-HA antibody (**I, Figure 2**). Subcellular location patterns of each MAC-tagged marker proteins were compared with the data from the Human Protein Atlas (https://www.proteinatlas.org/). Moreover, subcellular localizations of several MAC-tagged bait proteins (11/18 tested) were verified using specific antibodies against the corresponding endogenous protein (**I, Supplementary Figure 2**). In summary, MAC-tagged proteins localized to their corresponding well-documented cellular compartments, illustrating that the MAC-tag does not change the correct localization of the bait protein.

In the presence of biotin, biotinylated proteins could be detected with immunofluorescence or western (**I, Figure 2; I, Supplementary Figure 1 b; III, Figure 2 c**). These results suggest that the localization of the *in vivo* biotinylated proteins correlates well in close proximity with that of the corresponding bait proteins, indicating

the biotin signal generated by the MAC-tag constructs are coincident with the localization of bait proteins.

1.2 AP-MS and BioID are complementary approaches for the interactome mapping

We went on and applied AP-MS and BioID pipelines on 18 MAC-tagged cellular markers to obtain a comprehensive protein interaction information (**I, Supplementary Data 1a**). The detailed LC-MS set-up that we have used for the analysis was described in manuscript III. Computer approaches of significance analysis of interactome (SAINT) [227] and contaminant repository for affinity purification (CRAPome) [228] were used as statistical tools to assign a confidence score to the PPI based on appropriate control samples. In total, 679 HCIs were achieved by AP-MS, and 2118 HCIs from BioID were retained with the SAINT probabilistic score threshold of ≥ 0.73. **(Figure 17A)**.

To assess the complementary nature of these two methods, we illustrated the PPIs by their overlap as well as individual detection by two approaches **(Figure 17B)**. It is not surprising to notice the partial overlap between the two methods, considering the fact that BioID identifies the interactor of a bait protein in addition to its direct binding partners. In most cases, BioID generally produced larger interactomes than AP-MS, despite using the same cell lines, MS instrument, and statistical analysis (**I, Supplementary Figure 3**). Unlike AP-MS, the BioID approach is more sensitive in detecting transient and low-abundance interacting proteins. This feature is reflected by the percentage of identified novel interactions in total that is about two-fold higher using BioID (11.3) than that by AP-MS (6.8).

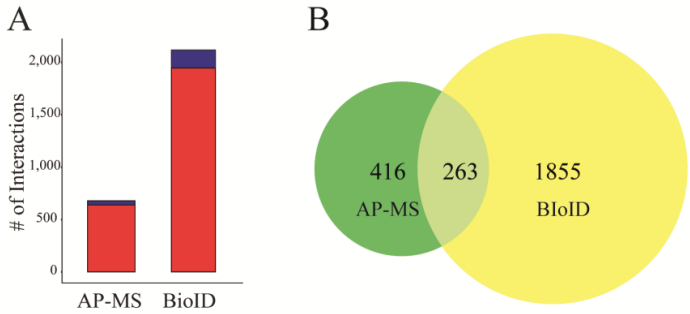


Figure 17. (A) The proportion of known (blue) and novel (red) interactions within 18 subcellular markers. (B) Venn's diagram shows the overlap of interactions between AP-MS (green) and BioID (yellow) methods. (Adopted and modified from Liu et al., 2018)

The complementary nature of AP-MS and BioID also reflects at the quantification level. By plotting the intensity data obtained from AP-MS and BioID, we could predict the relative spatial arrangement of the MAC-tagged bait protein to its interacting proteins in a stable protein complex (**I, Figure 7**). For example, MAC-tagged two components (CDK8 and MED13) of the mediator complex can be used to calculate the relative distance of the bait to the mediator core separately. The overall correlation of CDK8 and MED13 distances from the Mediator core is extremely high (c=0.95) (**I, Fig. 7b, c, g, h**). These results show that it is possible to derive the structural information of the protein complex by integrating AP-MS and BioID data. However, the calculation was based on the known stable protein complex. In future work, we plan to propose the protein structure information on a novel protein complex to direct the following EM study.

1.3 Mass spectrometry microscopy using the organelle-specific interactome as a reference

Proteins that can interact with each other tend to locate in the same compartments or physically adjacent areas, at least transiently or conditionally [229]. Based on such theory, interactomes of BirA* -tagged bait proteins locating in specific subcellular organelles can define the molecular context of that organelle. Thus, 2118 HCIs of the 18 *bona fide* localization markers obtained by the BioID approach formed an organelle-specific protein interaction network to serve as the reference grid. Overlaying PPIs data of any MAC-tagged (BioID approach) bait protein with the molecular reference map should allow defining the dynamic localization of the bait protein. We named this process as mass spectrometry microscopy (MS-microscopy) approach **(I, Figure 4a)**.

We validated the MS-microscopy approach using Golgi vesicular membrane-trafficking protein p18 (BET1), CDK7, GSK3B, and ras-related protein Rab-5A (RAB5A) as baits. BET1 participates in a late stage of ER-Golgi transport and is restricted mainly to the face of the Golgi apparatus [230] **(I, Figure 4b)**. CDK7 is mainly associating with chromatin in line with its essential role in transcription regulation **(I, Figure 4b)**. GSK3B is a major signaling node that mediates several signaling pathways [231]. MS-microscopy indicated GSK3B localizations to Golgi and exosome **(I, Figure 4b)**, which supports the recent finding that GSK3B locates in endosome with Wnt ligand and continues traveling with active Wnt through endosomal organelles onto exosomes [232].

Comparing with the method (Go, Knight et al. 2019) that annotates the prey localization based on the localization information of its baits, the MS-microscopy predicts the localization information of the bait based on the quantitative information of its preys. Thus, it is possible to provide more dynamic localization information not only to wild type bait protein but also to the genetic mutations in the context of human diseases [233, 234].

In summary, the application of MAC-tag approach on 18 cellular localization markers enables to obtain a comprehensive molecular context proteome map formed by these proteins, which can be used to build up the MS-microscopy system.

2 MAC-tag application on the mitochondrion

Previous results showed the MS-microscopy is capable of defining the molecular context of (m)any protein(s) at the subcellular level. In this chapter, we discuss the application of MAC-tag on the sub-organelle level and use mitochondrion as an example.

2.1 Sub-organelle level resolution of the MS-microscopy

Traditional confocal microscope with an optical resolution of 200-300nm, cannot distinguish whether a tagged protein is in the right mitochondrial sub-compartment. The intermembrane space (IMS) is 10-20 nm in diameter that is about the same distance as the labeling radius of BioID. Therefore, the interactors of MAC-tagged POIs locating in different sub-compartments of mitochondria should have unique proteome profiles. We went on to MAC-tag three mitochondrial proteins **(I, Supplementary Data 1a)** (PDK1 [matrix], SCO1 [IMS], and TOM20 [OMM]) and generated a sub-organelle molecular reference database of the mitochondrion **(I, Figure 6a)**. This mitochondrial sub-organelle molecular context consists of 121 (OMM), 102 (IMS), and 235 (matrix) proteins. To determine whether the mitochondrial reference database can be used as an indicator for the spatial arrangement of mitochondria proteome, we processed additional 13 mitochondrial proteins to define their mitochondrial sub-organelle localization **(I, Figure 6b)**. Our results agree with previous findings. Especially for aurora kinase A-interacting protein (AKIP), besides the mitochondrial matrix localization, the MS-microscopy also suggested AKIP has the chromatin localization within the nucleus **(I, Figure 6b)** where AKIP has been reported to be a function [235].

2.2 Interatomic insights of human GRPELs in the mitochondrial matrix

Most mitochondrial proteins are nuclear-encoded and are translocated from the cytosol into the matrix compartment. This process is facilitated by the mitochondrial heat shock protein 70 (mtHsp70) [236]. MtHsp70 interacts with the preprotein in transit in an ATP-dependent manner [237]. Previous studies have shown that GrpE-like protein 1 (GRPEL1) functions as a nucleotide exchange factor for mtHsp70 in human cells [238]. However, there is another GrpE-like protein, GRPEL2, existing in human cells. Despite function as co-chaperones for mtHsp70, their specific role in humans is still less well understood [239]. Therefore, to obtain a further comprehensive view of the function of GRPEL1 and GRPEL2, we implemented the BioID approach to reveal the interaction partners of GRPEL1 and GRPEL2 in human 143B cells.

To obtain the HCIs of GRPELs, BirA* tagged GFP, and apoptosis-inducing factor (AIF) were used as controls for SAINT scoring to generate a high-confidence candidate interacting protein (HCIP) list containing 44 candidate proteins **(Figure 18; II, Supplementary Table S1)**.

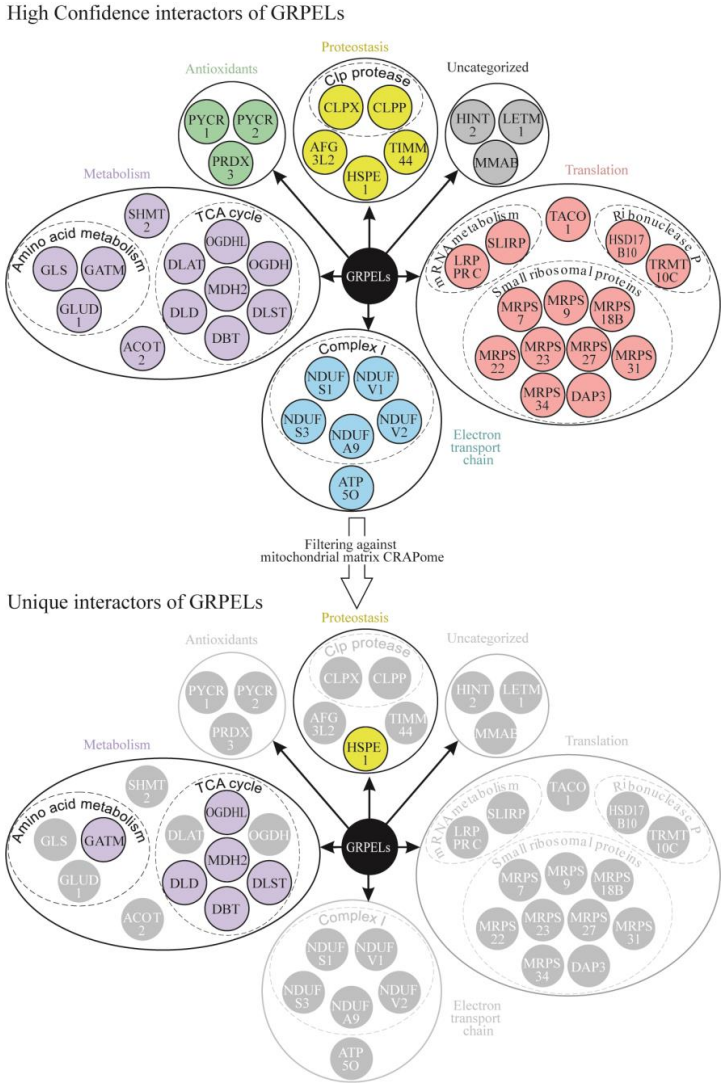


Figure 18. High confidence interactor of GRPEL. HCIs of GRPEL before (upper) and after filtering against mitochondrial-matrix-CRAPome (lower). The proteins filtered out with mitochondrial-matrix are colored by light gray in the lower panel. (Adapted and modified from Konovalova et al., 2018)

Although SAINT and CRAPome have proved to be useful tools for PPIs data filtering, the HCIs list still included many proteins that were ubiquitously identified as interactors for mitochondrial matrix protein studies, such as mitochondrial ribosome proteins. Therefore, we compared the results with interactomes of five MAC-tagged mitochondrial proteins (AKIP, CH60, MRM, PDK1, and TR61B) (**II, Supplementary Table S2**) to obtain a more accurate view of background protein for this specific sub-organelle. We excluded hits that are with a lower fold change of spectral count and higher frequency of occurrence from the HCIs list to generate a final HCIs list contains seven unique binding partners of GRPELs **(Figure 18)**. Interestingly, most of them were dehydrogenases. Further studies are required to investigate the function of GRPELs and dehydrogenase activities.

To sum up, the MAC-tag approach for interactome analysis not only can be used to verify the unique interaction partner of bait protein but also markedly expanded the MS-microscopy system to sub-organelle level.

3 MAC-tag application on Hh signaling pathway

Hh signaling pathway has a crucial role in several developmental processes, and its aberrant activation relates to a variety of cancers [92, 240]. Decoding the interactome of Hh signaling might reveal novel targets for cancer therapy [241, 242]. Thus, we applied MAC-tag on the well characterized, conserved canonical components of the pathway and used the BioID approach to establish the interaction network of Hh signaling.

3.1 Dynamics of interaction networks upon Hh pathway activation

To resolve the interactome underlying mammalian Hh signaling activation, 20 stable 3T3 cell lines (**IV, Supplementary Table 1a)** that constitutively express the conserved core components of the mammalian Hh signaling in endogenous level were generated to monitor the dynamic of PPI network in the absence and presence of Hh ligand. To confirm that the MAC-tagged bait proteins have correct localization and dynamic changes during pathway activation, we examined the stable cells by immunofluorescence (**IV, Figure 1**). The immunostaining results revealed that the majority of the MAC-tagged bait proteins were localize proximally to primary cilia or the regions approaching the cilia either on the signaling 'Off' or 'On' stage, indicating these isogenic stable cell lines (MAC-tagged POIs) can be activated upon Hh ligand treatment. We went on to proceed the MS analysis on 120 samples covering 19 bait proteins with three biological replicates in the presence/absence Hh ligand. A total of 3,569 HCIs were identified among 913 unique HCIPs with SAINT probabilistic score > 0.7 (Bayesian false discovery rate < 0.05) (**IV, Supplemenaty Table 1b,e**). More specially, 1681 HCIs were obtained in the absence of Hh ligand, and 1,888 HCIs were detected with pathway activation (**IV, Figure 2a**). Among them, 1,229 HCIs were consistently detected in both conditions (**IV, Figure 2a**). Some of the bait proteins were also recognized as prey proteins because of the interactions between the core components. To systematically characterize interactome of Hh signaling, we compared our HCIPs with the candidate gene list of the Hh signaling

pathway (P-value <0.05) that was obtained by CRISPR screening. (Pusapati, Kong et al. 2018). More than 28% of HCIPs (**Supplementary Figure 2a**) in both conditions were also considered as regulators or attenuators for Hh signaling in CRISPR screening [243], demonstrating the utility of our PPI network. The enrichment analysis of pathways based on gene ontology (GO) for both conditions showed that a large number of proteins were involved in the G1/S transition that regulates the primary cilia progression [244] without pathway activation. Interestingly, the non-canonical Wnt signaling was enriched in the absence of Hh, suggesting it may negatively regulate (downregulate) the Hh signaling [245, 246]. Not surprisingly, Hh stimulation significantly increased the portion of prey proteins involving in the Hh signaling pathway (**IV, Figure 2c**).

Zooming for details, the proteins EVC and WDR35 (WD repeat-containing protein 35) that form the complex with SMO during Hh pathway activation [247] were observed only under Hh ligand treatment condition (**IV, Supplementary Figure 1b**). Furthermore, using siRNA treatments on two HCIPs of SMO, neurobeachin (NBEA), and lipopolysaccharide-responsive and beige-like anchor protein (LRBA) can reduce Hh-activated GLI-luciferase expression in 3T3 cells (**IV, Figure 6d**). All these results indicate that the BioID approach of MAC-tag indeed is sensitive to detect the potential regulators of the Hh pathway.

3.2 Dynamics of interaction network during cilia progression

All the proteins must be transported into the primary cilium [104], there is significant interest in understanding whether protein interactions of these key components function inside the primary cilia or act independently. HEK293 cells were used for the studies as it has been used for constructing the protein landscape of the human primary cilia and centrosome-cilium interface [248]. Moreover, the interactome of core Hh pathway components (kinesin-like protein 7 (KIF7) [206], SUFU [249] and PKA [113] were obtained using HEK293 cells.

We selected 30 baits for BioID analysis in HEK293 cells. This bait set included the initial 19 core components of Hh signaling and another 11 cilia/Hh signaling relevant proteins **(IV, Supplementary Table 1a)**. In total, 35 MAC-tagged HEK293 stable cell lines were generated (**IV, Figure 3a; IV, Supplementary Table 1a**). We prepared the samples at specific points of cell cycle progressions to regulate the cilia dynamics. Therefore, the PPIs were monitored under the following conditions without Hh ligand exposure: 1). normal condition (cells were cultured in complete medium); 2). ciliogenesis condition (cells were cultured in serum-starved medium); and 3). ciliary resorption condition (cells were first induced to ciliogenesis by serum starvation and then induced to ciliary resorption by adding serum back to the culture medium) (**IV, Figure 3a**). Altogether, we performed 315 BioID purifications for 30 baits under three different biological conditions. The analysis resulted in 9,059 HCIs involving 1,600 HCIPs with SAINT probabilistic score >0.75 (Bayesian false discovery rate < 0.05) (**IV, Figure 3b; IV, Supplementary Table 1c**). On average, we identified ~86 HCIs for each bait, which

corresponds well to the number of interactions typically found in the published large-scale BioID studies [248, 250, 251]. Of the HCIPs, 26% were cilium related proteins (**IV, Figure 3d**), which is identical to what was earlier identified in NIH-3T3 cells (**IV, Figure 2b**). In total, 3,076 HCIs were identified under cilia formation, 2,845 HCIs from normal culture conditions, and 2,562 HCIs under resorption conditions (**IV, Figure 3 b,c**). GO analysis was conducted to elucidate the molecular roles of primary cilia on these PPIs, the most significant proportion of HCIPs associated with Hh signaling were obtained in ciliogenesis condition, highlighting the important role of cilia in Hh signaling. Though cilia are very critical for Hh signaling, we inevitably noticed that more than 50% of HCIs of each condition remained constant during the primary cilia progression (**IV, Figure 3c**). For example, the interaction between GLIs with SUFU was obtained in all three conditions (**IV, Figure 5a**), implying the fact that the regulation of GLI protein levels by SUFU is cilium and ligand-independent [111, 252]. The interactions of PTCH1 interacting with ubiquitin E3 ligase ITCH/WWP2 were also captured under all conditions (**IV, Figure 5a**), indicating that ubiquitylation regulates the degradation of PTCH1 independently from primary cilia [158].

3.3 MS-microscopy allows determination of protein localization in primary cilium

Unlike other cellular locations achieved by membrane-bound organelles, such as mitochondria and lysosomes, primary cilia are incompletely membrane-bounded and appear in the specific time of the cell cycle [253], making it difficult for localization mapping. ARL13B is highly enriched within primary cilia and has been extensively used as a marker to visualize primary cilia [254]. Thus, we incorporated the interaction profiles of ARL13B obtaining from HEK293 cells into the MS-microscopy reference database.

To test the applicability of the MS-microscopy on primary cilium, we applied it to all the 29 bait proteins that we used in cilia progression study. Under cilia formation and resorption conditions, PTCH1, PTCH2, SUFU, GLI1, BTCR, KANTB1, DHCR7, DLG5, and KIF3 are predominantly associating with primary cilia (**IV, Figure 4b**). Most of them, indeed, are well-documented ciliary proteins. Additionally, comparing with the cilia resorption condition, nine bait proteins show or reinforce the primary cilia localization in ciliogenesis conditions (**IV, Figure 4b**), suggesting their dynamic subcellular localization during cilia progression. These examples demonstrate the applicability of the MS-microscopy in defining protein localization in the primary cilium and verifying the correct subcellular localization of bait proteins in HEK293 cell lines.

To conclude, by monitoring dynamic changes of the interactome of core components in the presence/absence of Hh ligand and different cilia progression stages, we constitute the very first reported interaction landscape of the Hh pathway and provides a precious resource for directing future functional studies. Furthermore, we expanded the coverage of the MS-microscopy system to the primary cilium and benchmarked the usability of the MS-microscopy on transient organelle.

V CONCLUSION AND FUTURE PROSPECTS

A systematic investigation of mammalian PPI networks may provide important biological information for understanding the internal organization of a cell, uncovering the molecular mechanisms of signaling processes. In recent years, there have been some methods, including AP-MS and BioID, which have been successfully used for constructing and analyzing the protein interaction networks (chapter I). However, there is still a lack of a method that can be easily applied to identify and quantify protein complex stoichiometry and transient/proximal interactions. Therefore, this thesis project aims to develop a workflow that can integrate both AP-MS and BioID approaches to characterize the molecular context of POI.

Study I, we developed and optimized MAC-tag based workflow, which integrated Strep-tag for affinity purification and BirA* for BioID identification. Our workflow requires the generation of one cell line for switching from AP-MS to BioID by biotin addition. The single affinity matrix can be applied for both AP-MS and BioID during purification steps, improving the data reproducibility between two methods. AP-MS and BioID are complementary approaches to provide a complete interactome map of the POI. The integrated workflow greatly reduces the time for sample preparation and facilitates large-scale PPIs analysis. Moreover, we shared a detailed protocol (Study III) as part of the project for other researchers to follow.

Proteins form protein complexes and travel between different subcellular localizations to execute biological processes. The development of MS-based proteomics has provided new possibilities to query the spatial organization of the proteome on a larger scale (Chapter II). In study II, we implemented the BioID approach to reveal the interaction partners of GRPEL1 and GRPEL2 in human 143B cells. The analysis generated 44 candidate proteins. To further filter localization-based biologically unlikely interactors of GRPELs, we utilized the interactomes of MAC-tagged mitochondrial matrix proteins to exclude frequent detections in the BioID approach. Finally, we obtained the unique HCIP list containing seven interactors of GRPELs. The results facilitate the subsequent investigation about the function of GRPELs. In study I, we applied the MAC-tag approach on 18 cellular localization markers that are commonly used in immunofluorescence studies. The proximal interaction data obtained by BioID provided a molecular context for these cellular compartments. This protein context map can be used as an index to probe other proteins' subcellular localization by aligning the interactome data. We propose this mapping system as "MS–microscopy" (www.proteomics.fi) to localize POIs with high spatial resolution. Zooming in even further, we extended the resolution of MS-microscopy to a sub-organelle level. Three marker proteins of mitochondria were chosen to generate the mitochondrial reference proteome map. By applying it, we could establish the sub-mitochondrial localization of the POI.

Furthermore, in study IV, we expanded the MS-microscopy to a transient organelle, primary cilium, by incorporating interaction profiles of ARL13B into the MS-microscopy reference database. The updated version of MS-microscopy successfully verified the

correct ciliary localization of several Hh signaling related POIs. It worth to mention that our MS-microscopy is not only applicable to the MAC-tagged POI but also could be used for other BirA*-tagged POIs (Study I, Study III).

Over the past decades, Hh signaling has been extensively studied. However, much is still unknown about the signal transduction. The process of signal transduction is dependent on specific PPIs. Multiple small-scale studies have reported the protein interaction data by focusing on a few components in the Hh signaling pathway (Chapter III). In study IV, to elucidate the interactome involving in the mammalian Hh signaling, we took advantage of the MAC-tag system and followed by BioID purification in 3T3 cells to identify associated proteins of the Hh pathway. We selected 19 baits proteins that cover all the critical components of the Hh signaling to construct the Hh signaling pathway interactome in the absence and presence of Hh ligand. Altogether, we analyzed 120 samples and generated 3,556 HCIs. The PPIs establish the logical framework for our current understanding of the Hh signal transduction and assert the essence of primary cilia in Hh signaling. Meanwhile, to broaden our knowledge of the relationship between primary cilia and Hh signaling pathway, we chose 30 baits proteins that have been previously linked to the Hh signaling and ciliary structure, to analyze their interactome using HEK293 cells in contexts of normal, ciliogenesis and ciliary resorption conditions. In total, we identified 9,059 HCIs involving 1,600 HCIPs. The significant overlap of the interaction between three conditions indicated proteins could regulate the Hh signaling outside the primary cilia. This work provides valuable new insights into the function of PPIs in Hh signal transduction but also raises new questions. For example, are these interactions also involve in other signaling pathway activation? What will be the effects on the Hh signaling transduction by losing these interactions? Therefore, further research about the specific roles of these PPIs in the context of different tissues and cancer types will hopefully contribute to a better understanding of the Hh signaling pathway.

Taken together, this thesis presents the MAC-tag-based workflow for an in-depth analysis of protein interaction. The workflow can provide a comprehensive view of the protein interactome and transient signaling components. Moreover, MS-microscopy enables one to define the molecular location of the POI in both organelle and sub-organelle levels.

VI ACKNOWLEDGEMENTS

This thesis project was carried out in the Systems Biology/Pathology Research Group and Proteomic Unit at the Institute of Biotechnology, Helsinki Institute of Life Science (HILIFE), University of Helsinki. Financial support for the thesis work was obtained from the Doctoral Programme in Integrative Life Science.

My exceptional thanks go to my supervisor, Dr. Markku Varjosalo. Thank you for providing a constructive work environment and the best resources to pursue my scientific interests and try out things. Your enthusiasm, persistence, and encouragement in science paved the way for many experiments, even when I had doubts myself.

My thesis committee meetings were a valuable source of ideas and support throughout the project. I wish to thank Prof. Pekka Laplainen and Prof. Juha Partanen for their helpful suggestions and scientific impact for this project. My thesis pre-examiners Dr. Kirsi Rilla, and Dr. Pieta Mattila are thanked for their careful reviews of the thesis, and for all the constructive comments and suggestions that improved the thesis quality significantly. My sincere thanks to Doc. Franz Herzog for kindly agreeing to serve as an opponent for my thesis and being flexible with the defense date. I would also like to thank Prof. Ville Hietakangas for serving as Custos for my dissertation.

I wish to thank all of the co-authors and collaborators for their research contributions. Kari Salokas, Fitsum Tamene, Dr.Yaming Jiu, Rigbe Weldatsadik, Rigbe G. Weldatsadik, Dr.Tiina Öhman, and Lisa Gawriyski are thanked for their help with the MAC-tag project. Likewise, I am also grateful to Dr. Svetlana Konovalova and Prof. Henna Tyynismaa, for collaboration and for providing the opportunity to work with the mitochondrial system. Many thanks to Kari Salokas, Dr. Kubilay Demir, and Prof. Philip Beachy for supporting the Hedgehog project.

My warmest thanks to all fellows of the Varjosalo research group for their help over the years. Working with you: Salla, Sini, Leena, Helka, Kari, Jaakko, Matias, Tiina, Rigbae, Lisa, Lukas, Antti, Elina, and Saara, was indeed a pleasure. Special thanks to Salla, Sini, Helka, and Leena along the entire journey. This thesis could not have been achieved without their encouragement and advice.

Many thanks to my friends: Yunxian, Wangliang, Yixin, Liuying, Yaming, Hongxia, Changwei, Tahira, Dorota, and Adrian for all sorts of help, discussions, and other events.

My doctoral study has been a journey filled with ups and downs. I could not have survived without my wonderful family, who has been there every step of the way. I am truly grateful for all your support.

Vantaa, April 2020

Xiaonan Liu

VII REFERENCES

1. De Las Rivas, J. and C. Fontanillo, *Protein-protein interactions essentials: key concepts to building and analyzing interactome networks.* PLoS Comput Biol, 2010. **6**(6): p. e1000807.
2. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.
3. Perkins, J.R., et al., *Transient Protein-Protein Interactions: Structural, Functional, and Network Properties.* Structure, 2010. **18**(10): p. 1233-1243.
4. Eisenberg, D., et al., *Protein function in the post-genomic era.* Nature, 2000. **405**(6788): p. 823-6.
5. Snider, J., et al., *Fundamentals of protein interaction network mapping.* Mol Syst Biol, 2015. **11**(12): p. 848.
6. Hao, T., et al., *Reconstruction and Application of Protein-Protein Interaction Network.* Int J Mol Sci, 2016. **17**(6).
7. Pan, S., et al., *Mass spectrometry based targeted protein quantification: methods and applications.* J Proteome Res, 2009. **8**(2): p. 787-97.
8. Rauh, M., *LC-MS/MS for protein and peptide quantification in clinical chemistry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2012. **883-884**: p. 59-67.
9. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
10. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annual review of biomedical engineering, 2009. **11**: p. 49-79.
11. Bordas-Nagy, J., D. Despeyroux, and K.R. Jennings, *Comparison of helium and argon as collision gases in the high energy collision-induced decomposition of MH(+) ions of peptides.* J Am Soc Mass Spectrom, 1992. **3**(5): p. 502-14.
12. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
13. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
14. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
15. Higdon, R. and E. Kolker, *A predictive model for identifying proteins by a single peptide match.* Bioinformatics, 2006. **23**(3): p. 277-280.
16. Gingras, A.C., et al., *Analysis of protein complexes using mass spectrometry.* Nat Rev Mol Cell Biol, 2007. **8**(8): p. 645-54.
17. Varjosalo, M., et al., *Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS.* Nat Methods, 2013. **10**(4): p. 307-14.
18. Dunham, W.H., M. Mullin, and A.C. Gingras, *Affinity-purification coupled to mass spectrometry: basic principles and strategies.* Proteomics, 2012. **12**(10): p. 1576-90.
19. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.
20. Duplay, P., et al., *Sequences of the malE gene and of its product, the maltose-binding protein of Escherichia coli K12.* J Biol Chem, 1984. **259**(16): p. 10606-13.
21. Stretton, S., et al., *Use of green fluorescent protein to tag and investigate gene expression in marine bacteria.* Appl Environ Microbiol, 1998. **64**(7): p. 2554-9.
22. Hubner, N.C., et al., *Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions.* J Cell Biol, 2010. **189**(4): p. 739-54.
23. Hein, M.Y., et al., *A human interactome in three quantitative dimensions organized by stoichiometries and abundances.* Cell, 2015. **163**(3): p. 712-23.
24. Sun, P., J.E. Tropea, and D.S. Waugh, *Enhancing the solubility of recombinant proteins in Escherichia coli by using hexahistidine-tagged maltose-binding protein as a fusion partner.* Methods Mol Biol, 2011. **705**: p. 259-74.

25. Bokhove, M., et al., *Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins.* J Struct Biol, 2016. **194**(1): p. 1-7.
26. Reuten, R., et al., *Maltose-Binding Protein (MBP), a Secretion-Enhancing Tag for Mammalian Protein Expression Systems.* PLoS One, 2016. **11**(3): p. e0152386.
27. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.
28. Evan, G.I., et al., *Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product.* Mol Cell Biol, 1985. **5**(12): p. 3610-6.
29. Hopp, T.P., et al., *A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification.* Bio/Technology, 1988. **6**(10): p. 1204-1210.
30. Schmidt, T.G. and A. Skerra, *The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment.* Protein Eng, 1993. **6**(1): p. 109-22.
31. Slootstra, J.W., et al., *Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5.* 1997. **2**(3): p. 156-164.
32. Simsek, D., et al., *The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity.* Cell, 2017. **169**(6): p. 1051-1065.e18.
33. Schmidt, T. and A. Skerra, *The Strep-tag system for one-step affinity purification of proteins from mammalian cell culture.* Methods Mol Biol, 2015. **1286**: p. 83-95.
34. Wilchek, M. and E.A. Bayer, *Introduction to avidin-biotin technology.* Methods Enzymol, 1990. **184**: p. 5-13.
35. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
36. Varjosalo, M., et al., *The protein interaction landscape of the human CMGC kinase group.* Cell Rep, 2013. **3**(4): p. 1306-20.
37. Bonetta, L., *Protein-protein interactions: Interactome under construction.* Nature, 2010. **468**(7325): p. 851-4.
38. Roux, K.J., et al., *A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells.* J Cell Biol, 2012. **196**(6): p. 801-10.
39. Choi-Rhee, E., H. Schulman, and J.E. Cronan, *Promiscuous protein biotinylation by Escherichia coli biotin protein ligase.* Protein Sci, 2004. **13**(11): p. 3043-50.
40. Kim, D.I., et al., *Probing nuclear pore complex architecture with proximity-dependent biotinylation.* Proc Natl Acad Sci U S A, 2014. **111**(24): p. E2453-61.
41. Kim, D.I., S.C. Jensen, and K.J. Roux, *Identifying Protein-Protein Associations at the Nuclear Envelope with BioID.* Methods Mol Biol, 2016. **1411**: p. 133-46.
42. Uezu, A., et al., *Identification of an elaborate complex mediating postsynaptic inhibition.* Science, 2016. **353**(6304): p. 1123-9.
43. Ramanathan, M., et al., *RNA-protein interaction detection in living cells.* Nat Methods, 2018. **15**(3): p. 207-212.
44. Chojnowski, A., et al., *2C-BioID: An Advanced Two Component BioID System for Precision Mapping of Protein Interactomes.* iScience, 2018. **10**: p. 40-52.
45. Kim, D.I., et al., *An improved smaller biotin ligase for BioID proximity labeling.* Mol Biol Cell, 2016. **27**(8): p. 1188-96.
46. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.
47. Putyrski, M. and C. Schultz, *Protein translocation as a tool: The current rapamycin story.* FEBS Lett, 2012. **586**(15): p. 2097-105.
48. De Munter, S., et al., *Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions.* FEBS Lett, 2017. **591**(2): p. 415-424.
49. Schopp, I.M., et al., *Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes.* Nat Commun, 2017. **8**: p. 15690.

50. Martell, J.D., et al., *Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy.* Nat Biotechnol, 2012. **30**(11): p. 1143-8.
51. Rhee, H.W., et al., *Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging.* Science, 2013. **339**(6125): p. 1328-1331.
52. Lam, S.S., et al., *Directed evolution of APEX2 for electron microscopy and proximity labeling.* Nat Methods, 2015. **12**(1): p. 51-4.
53. Han, Y., et al., *Directed Evolution of Split APEX2 Peroxidase.* ACS Chem Biol, 2019. **14**(4): p. 619-635.
54. Li, X.W., et al., *New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay.* J Biol Chem, 2014. **289**(21): p. 14434-47.
55. Rees, J.S., et al., *Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters.* Curr Protoc Protein Sci, 2015. **80**: p. 19.27.1-19.27.18.
56. Bar, D.Z., et al., *Biotinylation by antibody recognition-a method for proximity labeling.* Nat Methods, 2018. **15**(2): p. 127-133.
57. Liu, Q., et al., *A proximity-tagging system to identify membrane protein-protein interactions.* Nat Methods, 2018. **15**(9): p. 715-722.
58. Zhuang, M., et al., *Substrates of IAP ubiquitin ligases identified with a designed orthogonal E3 ligase, the NEDDylator.* Mol Cell, 2013. **49**(2): p. 273-82.
59. Young, M.M., et al., *High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5802-6.
60. Sinz, A., *Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions.* Mass Spectrom Rev, 2006. **25**(4): p. 663-82.
61. Back, J.W., et al., *Chemical cross-linking and mass spectrometry for protein structural modeling.* J Mol Biol, 2003. **331**(2): p. 303-13.
62. Green, N.S., E. Reisler, and K.N. Houk, *Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers.* Protein Sci, 2001. **10**(7): p. 1293-304.
63. Bruce, J.E., *In vivo protein complex topologies: sights through a cross-linking lens.* Proteomics, 2012. **12**(10): p. 1565-75.
64. Chavez, J.D., et al., *Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus-plant interactions.* J Proteome Res, 2012. **11**(5): p. 2968-81.
65. Navare, A.T., et al., *Probing the protein interaction network of Pseudomonas aeruginosa cells by chemical cross-linking mass spectrometry.* Structure, 2015. **23**(4): p. 762-73.
66. Chavez, J.D., et al., *In Vivo Conformational Dynamics of Hsp90 and Its Interactors.* Cell Chem Biol, 2016. **23**(6): p. 716-26.
67. Chavez, J.D., et al., *Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue.* Cell Syst, 2018. **6**(1): p. 136-141.e5.
68. Chavez, J.D., et al., *Systems structural biology measurements by in vivo cross-linking with mass spectrometry.* Nat Protoc, 2019. **14**(8): p. 2318-2343.
69. Eyckerman, S., et al., *Trapping mammalian protein complexes in viral particles.* Nat Commun, 2016. **7**: p. 11416.
70. Lundberg, E. and G.H.H. Borner, *Spatial proteomics: a powerful discovery tool for cell biology.* Nat Rev Mol Cell Biol, 2019. **20**(5): p. 285-302.
71. Kang, R., et al., *The Beclin 1 network regulates autophagy and apoptosis.* Cell Death Differ, 2011. **18**(4): p. 571-80.
72. Shimizu, S., M. Narita, and Y. Tsujimoto, *Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC.* Nature, 1999. **399**(6735): p. 483-7.
73. Popgeorgiev, N., L. Jabbour, and G. Gillet, *Subcellular Localization and Dynamics of the Bcl-2 Family of Proteins.* Front Cell Dev Biol, 2018. **6**: p. 13.

74. Roix, J. and T. Misteli, *Genomes, proteomes, and dynamic networks in the cell nucleus.* Histochem Cell Biol, 2002. **118**(2): p. 105-16.
75. Fasci, D., et al., *Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei.* Mol Cell Proteomics, 2018. **17**(10): p. 2018-2033.
76. Perier, C. and M. Vila, *Mitochondrial biology and Parkinson's disease.* Cold Spring Harb Perspect Med, 2012. **2**(2): p. a009332.
77. Azimzadeh, P., et al., *Comparison of three methods for mitochondria isolation from the human liver cell line (HepG2).* Gastroenterol Hepatol Bed Bench, 2016. **9**(2): p. 105-13.
78. Liu, F., et al., *The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes.* Mol Cell Proteomics, 2018. **17**(2): p. 216-232.
79. Yoon, J., et al., *Revealing Nanoscale Morphology of the Primary Cilium Using Super-Resolution Fluorescence Microscopy.* Biophys J, 2019. **116**(2): p. 319-329.
80. Mick, D.U., et al., *Proteomics of Primary Cilia by Proximity Labeling.* Dev Cell, 2015. **35**(4): p. 497-512.
81. Kohli, P., et al., *The ciliary membrane-associated proteome reveals actin-binding proteins as key components of cilia.* EMBO Rep, 2017. **18**(9): p. 1521-1535.
82. Boldt, K., et al., *An organelle-specific protein landscape identifies novel diseases and molecular mechanisms.* Nat Commun, 2016. **7**: p. 11491.
83. Kwon, Y., et al., *The Hippo signaling pathway interactome.* Science, 2013. **342**(6159): p. 737-40.
84. Couzens, A.L., et al., *Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions.* Sci Signal, 2013. **6**(302): p. rs15.
85. Hauri, S., et al., *Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1.* Mol Syst Biol, 2013. **9**: p. 713.
86. Wang, W., et al., *Defining the protein-protein interaction network of the human hippo pathway.* Mol Cell Proteomics, 2014. **13**(1): p. 119-31.
87. Nusslein-Volhard, C. and E. Wieschaus, *Mutations affecting segment number and polarity in Drosophila.* Nature, 1980. **287**(5785): p. 795-801.
88. Varjosalo, M., S.P. Li, and J. Taipale, *Divergence of hedgehog signal transduction mechanism between Drosophila and mammals.* Dev Cell, 2006. **10**(2): p. 177-86.
89. Ryan, K.E. and C. Chiang, *Hedgehog secretion and signal transduction in vertebrates.* J Biol Chem, 2012. **287**(22): p. 17905-13.
90. Li, J., et al., *PKA-mediated Gli2 and Gli3 phosphorylation is inhibited by Hedgehog signaling in cilia and reduced in Talpid3 mutant.* Dev Biol, 2017. **429**(1): p. 147-157.
91. Humke, E.W., et al., *The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins.* Genes Dev, 2010. **24**(7): p. 670-82.
92. Varjosalo, M. and J. Taipale, *Hedgehog: functions and mechanisms.* Genes Dev, 2008. **22**(18): p. 2454-72.
93. Fabregat, A., et al., *The Reactome Pathway Knowledgebase.* Nucleic Acids Res, 2018. **46**(D1): p. D649-d655.
94. Sidiropoulos, K., et al., *Reactome enhanced pathway visualization.* Bioinformatics, 2017. **33**(21): p. 3461-3467.
95. Wang, D., et al., *Aberrant activation of hedgehog signaling promotes cell proliferation via the transcriptional activation of forkhead Box M1 in colorectal cancer cells.* J Exp Clin Cancer Res, 2017. **36**(1): p. 23.
96. Scales, S.J. and F.J. de Sauvage, *Mechanisms of Hedgehog pathway activation in cancer and implications for therapy.* Trends Pharmacol Sci, 2009. **30**(6): p. 303-12.
97. Brennan, D., et al., *Noncanonical Hedgehog signaling.* Vitam Horm, 2012. **88**: p. 55-72.
98. Gu, D. and J. Xie, *Non-Canonical Hh Signaling in Cancer-Current Understanding and Future Directions.* Cancers (Basel), 2015. **7**(3): p. 1684-98.
99. Szczepny, A., et al., *The role of canonical and non-canonical Hedgehog signaling in tumor progression in a mouse model of small cell lung cancer.* Oncogene, 2017. **36**(39): p. 5544-5550.

100. Pietrobono, S., S. Gagliardi, and B. Stecca, *Non-canonical Hedgehog Signaling Pathway in Cancer: Activation of GLI Transcription Factors Beyond Smoothened.* Front Genet, 2019. **10**: p. 556.
101. Brechbiel, J., K. Miller-Moslin, and A.A. Adjei, *Crosstalk between hedgehog and other signaling pathways as a basis for combination therapies in cancer.* Cancer Treat Rev, 2014. **40**(6): p. 750-9.
102. Bangs, F. and K.V. Anderson, *Primary Cilia and Mammalian Hedgehog Signaling.* Cold Spring Harb Perspect Biol, 2017. **9**(5).
103. Huangfu, D., et al., *Hedgehog signalling in the mouse requires intraflagellar transport proteins.* Nature, 2003. **426**(6962): p. 83-7.
104. Hsiao, Y.C., K. Tuz, and R.J. Ferland, *Trafficking in and to the primary cilium.* Cilia, 2012. **1**(1): p. 4.
105. Lechtreck, K.F., *IFT-Cargo Interactions and Protein Transport in Cilia.* Trends Biochem Sci, 2015. **40**(12): p. 765-778.
106. Rohatgi, R., L. Milenkovic, and M.P. Scott, *Patched1 Regulates Hedgehog Signaling at the Primary Cilium.* 2007. **317**(5836): p. 372-376.
107. Kim, J., et al., *The role of ciliary trafficking in Hedgehog receptor signaling.* Sci Signal, 2015. **8**(379): p. ra55.
108. Milenkovic, L., et al., *Single-molecule imaging of Hedgehog pathway protein Smoothened in primary cilia reveals binding events regulated by Patched1.* Proc Natl Acad Sci U S A, 2015. **112**(27): p. 8320-5.
109. Milenkovic, L., M.P. Scott, and R. Rohatgi, *Lateral transport of Smoothened from the plasma membrane to the membrane of the cilium.* J Cell Biol, 2009. **187**(3): p. 365-74.
110. Wang, Y., et al., *Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation.* 2009. **106**(8): p. 2623-2628.
111. Chen, M.H., et al., *Cilium-independent regulation of Gli protein function by Sufu in Hedgehog signaling is evolutionarily conserved.* Genes Dev, 2009. **23**(16): p. 1910-28.
112. Haycraft, C.J., et al., *Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function.* PLoS Genet, 2005. **1**(4): p. e53.
113. Chen, Y., et al., *Dual Phosphorylation of suppressor of fused (Sufu) by PKA and GSK3beta regulates its stability and localization in the primary cilium.* J Biol Chem, 2011. **286**(15): p. 13502-11.
114. Yao, E. and P.T. Chuang, *Hedgehog signaling: From basic research to clinical applications.* J Formos Med Assoc, 2015. **114**(7): p. 569-76.
115. He, M., et al., *The kinesin-4 protein Kif7 regulates mammalian Hedgehog signalling by organizing the cilium tip compartment.* Nat Cell Biol, 2014. **16**(7): p. 663-72.
116. Rosenbaum, J.L. and G.B. Witman, *Intraflagellar transport.* Nature Reviews Molecular Cell Biology, 2002. **3**(11): p. 813-825.
117. Mukhopadhyay, S., et al., *TULP3 bridges the IFT-A complex and membrane phosphoinositides to promote trafficking of G protein-coupled receptors into primary cilia.* Genes Dev, 2010. **24**(19): p. 2180-93.
118. Avidor-Reiss, T. and M.R. Leroux, *Shared and Distinct Mechanisms of Compartmentalized and Cytosolic Ciliogenesis.* Curr Biol, 2015. **25**(23): p. R1143-50.
119. Bujakowska, K.M., et al., *Mutations in IFT172 cause isolated retinal degeneration and Bardet-Biedl syndrome.* Hum Mol Genet, 2015. **24**(1): p. 230-42.
120. Chekuri, A., et al., *IFT88 mutations identified in individuals with non-syndromic recessive retinal degeneration result in abnormal ciliogenesis.* Hum Genet, 2018. **137**(6-7): p. 447-458.
121. Liu, A., B. Wang, and L.A. Niswander, *Mouse intraflagellar transport proteins regulate both the activator and repressor functions of Gli transcription factors.* Development, 2005. **132**(13): p. 3103-11.
122. Houde, C., et al., *Hippi is essential for node cilia assembly and Sonic hedgehog signaling.* Dev Biol, 2006. **300**(2): p. 523-33.

123. Schmidts, M., et al., *Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement.* J Med Genet, 2013. **50**(5): p. 309-23.
124. Ocbina, P.J. and K.V. Anderson, *Intraflagellar transport, cilia, and mammalian Hedgehog signaling: analysis in mouse embryonic fibroblasts.* Dev Dyn, 2008. **237**(8): p. 2030-8.
125. Ocbina, P.J., et al., *Complex interactions between genes controlling trafficking in primary cilia.* Nat Genet, 2011. **43**(6): p. 547-53.
126. Cantagrel, V., et al., *Mutations in the cilia gene ARL13B lead to the classical form of Joubert syndrome.* Am J Hum Genet, 2008. **83**(2): p. 170-9.
127. Bay, S.N., A.B. Long, and T. Caspary, *Disruption of the ciliary GTPase Arl13b suppresses Sonic hedgehog overactivation and inhibits medulloblastoma formation.* Proc Natl Acad Sci U S A, 2018. **115**(7): p. 1570-1575.
128. Legue, E. and K.F. Liem, Jr., *Tulp3 Is a Ciliary Trafficking Gene that Regulates Polycystic Kidney Disease.* Curr Biol, 2019. **29**(5): p. 803-812.e5.
129. Norman, R.X., et al., *Tubby-like protein 3 (TULP3) regulates patterning in the mouse embryo through inhibition of Hedgehog signaling.* Hum Mol Genet, 2009. **18**(10): p. 1740-54.
130. Caparros-Martin, J.A., et al., *The ciliary Evc/Evc2 complex interacts with Smo and controls Hedgehog pathway activity in chondrocytes by regulating Sufu/Gli3 dissociation and Gli3 trafficking in primary cilia.* Hum Mol Genet, 2013. **22**(1): p. 124-39.
131. Blair, H.J., et al., *Evc2 is a positive modulator of Hedgehog signalling that interacts with Evc at the cilia membrane and is also found in the nucleus.* BMC Biol, 2011. **9**: p. 14.
132. Mann, R.K. and P.A. Beachy, *Novel lipid modifications of secreted protein signals.* Annu Rev Biochem, 2004. **73**: p. 891-923.
133. Tukachinsky, H., et al., *Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand.* Cell Rep, 2012. **2**(2): p. 308-20.
134. Beachy, P.A., et al., *Interactions between Hedgehog proteins and their binding partners come into view.* Genes Dev, 2010. **24**(18): p. 2001-12.
135. Guo, W. and H. Roelink, *Loss of the Heparan Sulfate Proteoglycan Glypican5 Facilitates Long-Range Sonic Hedgehog Signaling.* Stem Cells, 2019. **37**(7): p. 899-909.
136. Chuang, P.T. and A.P. McMahon, *Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein.* Nature, 1999. **397**(6720): p. 617-21.
137. Tenzen, T., et al., *The Cell Surface Membrane Proteins Cdo and Boc Are Components and Targets of the Hedgehog Signaling Pathway and Feedback Network in Mice.* Developmental Cell, 2006. **10**(5): p. 647-656.
138. Kavran, J.M., et al., *All mammalian Hedgehog proteins interact with cell adhesion molecule, down-regulated by oncogenes (CDO) and brother of CDO (BOC) in a conserved manner.* J Biol Chem, 2010. **285**(32): p. 24584-90.
139. Gallardo, V. and P. Bovolenta, *Positive and negative regulation of Shh signalling in vertebrate retinal development.* F1000Res, 2018. **7**.
140. Izzi, L., et al., *Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation.* Dev Cell, 2011. **20**(6): p. 788-801.
141. Lee, C.S., L. Buttitta, and C.-M. Fan, *Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite.* 2001. **98**(20): p. 11347-11352.
142. Martinelli, D.C. and C.M. Fan, *Gas1 extends the range of Hedgehog action by facilitating its signaling.* Genes Dev, 2007. **21**(10): p. 1231-43.
143. Carpenter, D., et al., *Characterization of two patched receptors for the vertebrate hedgehog protein family.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13630-4.
144. Lu, X., S. Liu, and T.B. Kornberg, *The C-terminal tail of the Hedgehog receptor Patched regulates both localization and turnover.* Genes Dev, 2006. **20**(18): p. 2539-51.
145. Hsia, E.Y., Y. Gui, and X. Zheng, *Regulation of Hedgehog signaling by ubiquitination.* Front Biol (Beijing), 2015. **10**(3): p. 203-220.

146. Zhulyn, O., et al., *Ptch2 shares overlapping functions with Ptch1 in Smo regulation and limb development.* Dev Biol, 2015. **397**(2): p. 191-202.
147. Rahnama, F., R. Toftgard, and P.G. Zaphiropoulos, *Distinct roles of PTCH2 splice variants in Hedgehog signalling.* Biochem J, 2004. **378**(Pt 2): p. 325-34.
148. Gong, X., et al., *Structural basis for the recognition of Sonic Hedgehog by human Patched1.* Science, 2018. **361**(6402).
149. Smyth, I., et al., *Isolation and characterization of human patched 2 (PTCH2), a putative tumour suppressor gene inbasal cell carcinoma and medulloblastoma on chromosome 1p32.* Hum Mol Genet, 1999. **8**(2): p. 291-7.
150. Pusapati, G.V. and R. Rohatgi, *Location, location, and location: compartmentalization of Hedgehog signaling at primary cilia.* Embo j, 2014. **33**(17): p. 1852-4.
151. Holtz, A.M., et al., *Essential role for ligand-dependent feedback antagonism of vertebrate hedgehog signaling by PTCH1, PTCH2 and HHIP1 during neural patterning.* Development, 2013. **140**(16): p. 3423-34.
152. Qi, X., et al., *Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex.* Science, 2018. **362**(6410).
153. Qian, H., et al., *Inhibition of tetrameric Patched1 by Sonic Hedgehog through an asymmetric paradigm.* Nat Commun, 2019. **10**(1): p. 2320.
154. Bae, G.U., et al., *Mutations in CDON, encoding a hedgehog receptor, result in holoprosencephaly and defective interactions with other hedgehog receptors.* Am J Hum Genet, 2011. **89**(2): p. 231-40.
155. Izzi, L., et al., *Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation.* Dev Cell, 2011. **20**(6): p. 788-801.
156. Incardona, J.P., et al., *Receptor-mediated endocytosis of soluble and membrane-tethered Sonic hedgehog by Patched-1.* Proc Natl Acad Sci U S A, 2000. **97**(22): p. 12044-9.
157. Yue, S., et al., *Requirement of Smurf-mediated endocytosis of Patched1 in sonic hedgehog signal reception.* Elife, 2014. **3**.
158. Chen, X.L., et al., *Patched-1 proapoptotic activity is downregulated by modification of K1413 by the E3 ubiquitin-protein ligase Itchy homolog.* Mol Cell Biol, 2014. **34**(20): p. 3855-66.
159. Chen, Y. and G. Struhl, *In vivo evidence that Patched and Smoothened constitute distinct binding and transducing components of a Hedgehog receptor complex.* Development, 1998. **125**(24): p. 4943-8.
160. Mukhopadhyay, S. and R. Rohatgi, *G-protein-coupled receptors, Hedgehog signaling and primary cilia.* Semin Cell Dev Biol, 2014. **33**: p. 63-72.
161. Byrne, E.F., et al., *Multiple ligand binding sites regulate the Hedgehog signal transducer Smoothened in vertebrates.* Curr Opin Cell Biol, 2018. **51**: p. 81-88.
162. Thibert, C., et al., *Inhibition of neuroepithelial patched-induced apoptosis by sonic hedgehog.* Science, 2003. **301**(5634): p. 843-6.
163. Mille, F., et al., *The Patched dependence receptor triggers apoptosis through a DRAL-caspase-9 complex.* Nat Cell Biol, 2009. **11**(6): p. 739-46.
164. Fombonne, J., et al., *Patched dependence receptor triggers apoptosis through ubiquitination of caspase-9.* Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10510-5.
165. Infante, P., et al., *Targeting Hedgehog Signalling through the Ubiquitylation Process: The Multiple Roles of the HECT-E3 Ligase Itch.* Cells, 2019. **8**(2).
166. Katic, J., et al., *The cell adhesion molecule CHL1 interacts with patched-1 to regulate apoptosis during postnatal cerebellar development.* J Cell Sci, 2017. **130**(15): p. 2606-2619.
167. Barnes, E.A., et al., *Patched1 interacts with cyclin B1 to regulate cell cycle progression.* Embo j, 2001. **20**(9): p. 2214-23.
168. Jiang, X., P. Yang, and L. Ma, *Kinase activity-independent regulation of cyclin pathway by GRK2 is essential for zebrafish early development.* Proc Natl Acad Sci U S A, 2009. **106**(25): p. 10183-8.
169. Byrne, E.F.X., et al., *Structural basis of Smoothened regulation by its extracellular domains.* Nature, 2016. **535**(7613): p. 517-522.

170. Zhang, X., et al., *Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand.* Nat Commun, 2017. **8**: p. 15383.
171. Chen, Y., et al., *Sonic Hedgehog dependent phosphorylation by CK1α and GRK2 is required for ciliary accumulation and activation of smoothened.* PLoS Biol, 2011. **9**(6): p. e1001083.
172. Chong, Y.C., et al., *Bifurcating action of Smoothened in Hedgehog signaling is mediated by Dlg5.* Genes Dev, 2015. **29**(3): p. 262-76.
173. Chen, Y., et al., *Sonic Hedgehog dependent phosphorylation by CK1alpha and GRK2 is required for ciliary accumulation and activation of smoothened.* PLoS Biol, 2011. **9**(6): p. e1001083.
174. Xia, R., et al., *USP8 promotes smoothened signaling by preventing its ubiquitination and changing its subcellular localization.* PLoS Biol, 2012. **10**(1): p. e1001238.
175. Kovacs, J.J., et al., *Beta-arrestin-mediated localization of smoothened to the primary cilium.* Science, 2008. **320**(5884): p. 1777-81.
176. Barakat, B., et al., *Interaction of smoothened with integrin-linked kinase in primary cilia mediates Hedgehog signalling.* EMBO Rep, 2013. **14**(9): p. 837-44.
177. Yang, C., et al., *Smoothened transduces Hedgehog signal by forming a complex with Evc/Evc2.* Cell Res, 2012. **22**(11): p. 1593-604.
178. Dorn, K.V., C.E. Hughes, and R. Rohatgi, *A Smoothened-Evc2 complex transduces the Hedgehog signal at primary cilia.* Dev Cell, 2012. **23**(4): p. 823-35.
179. Pusapati, G.V., et al., *EFCAB7 and IQCE regulate hedgehog signaling by tethering the EVC-EVC2 complex to the base of primary cilia.* Dev Cell, 2014. **28**(5): p. 483-96.
180. Goetz, S.C., P.J. Ocbina, and K.V. Anderson, *The primary cilium as a Hedgehog signal transduction machine.* Methods Cell Biol, 2009. **94**: p. 199-222.
181. Polizio, A.H., et al., *Sonic Hedgehog activates the GTPases Rac1 and RhoA in a Gli-independent manner through coupling of smoothened to Gi proteins.* Sci Signal, 2011. **4**(200): p. pt7.
182. Ogden, S.K., et al., *G protein Galphai functions immediately downstream of Smoothened in Hedgehog signalling.* Nature, 2008. **456**(7224): p. 967-70.
183. Riobo, N.A., et al., *Activation of heterotrimeric G proteins by Smoothened.* Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12607-12.
184. Polizio, A.H., et al., *Heterotrimeric Gi proteins link Hedgehog signaling to activation of Rho small GTPases to promote fibroblast migration.* J Biol Chem, 2011. **286**(22): p. 19589-96.
185. Belgacem, Y.H. and L.N. Borodinsky, *Sonic hedgehog signaling is decoded by calcium spike activity in the developing spinal cord.* Proc Natl Acad Sci U S A, 2011. **108**(11): p. 4482-7.
186. Qu, C., et al., *Trimeric G protein-CARMA1 axis links smoothened, the hedgehog receptor transducer, to NF-kappaB activation in diffuse large B-cell lymphoma.* Blood, 2013. **121**(23): p. 4718-28.
187. Ruat, M., et al., *Targeting of Smoothened for therapeutic gain.* Trends Pharmacol Sci, 2014. **35**(5): p. 237-46.
188. Wu, F., et al., *Hedgehog Signaling: From Basic Biology to Cancer Therapy.* Cell Chem Biol, 2017. **24**(3): p. 252-280.
189. Luchetti, G., et al., *Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling.* Elife, 2016. **5**.
190. Huang, P., et al., *Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling.* Cell, 2016. **166**(5): p. 1176-1187.e14.
191. Zhang, Y., et al., *Structural insight into the mutual recognition and regulation between Suppressor of Fused and Gli/Ci.* Nat Commun, 2013. **4**: p. 2608.
192. Cherry, A.L., et al., *Structural basis of SUFU-GLI interaction in human Hedgehog signalling regulation.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 12): p. 2563-79.
193. Zhou, F., et al., *Nek2A/SuFu feedback loop regulates Gli-mediated Hedgehog signaling pathway.* Int J Oncol, 2017. **50**(2): p. 373-380.
194. Lin, C., et al., *Regulation of Sufu activity by p66beta and Mycbp provides new insight into vertebrate Hedgehog signaling.* Genes Dev, 2014. **28**(22): p. 2547-63.

195. Cheng, S.Y. and J.M. Bishop, *Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5442-7.
196. Paces-Fessy, M., et al., *The negative regulator of Gli, Suppressor of fused (Sufu), interacts with SAP18, Galectin3 and other nuclear proteins.* Biochem J, 2004. **378**(Pt 2): p. 353-62.
197. Infante, P., et al., *Itch/beta-arrestin2-dependent non-proteolytic ubiquitylation of SuFu controls Hedgehog signalling and medulloblastoma tumorigenesis.* Nat Commun, 2018. **9**(1): p. 976.
198. Raducu, M., et al., *SCF (Fbxl17) ubiquitylation of Sufu regulates Hedgehog signaling and medulloblastoma development.* Embo j, 2016. **35**(13): p. 1400-16.
199. Kim, Y.S., et al., *GLIS3, a novel member of the GLIS subfamily of Kruppel-like zinc finger proteins with repressor and activation functions.* Nucleic Acids Res, 2003. **31**(19): p. 5513-25.
200. ZeRuth, G.T., X.P. Yang, and A.M. Jetten, *Modulation of the transactivation function and stability of Kruppel-like zinc finger protein Gli-similar 3 (Glis3) by Suppressor of Fused.* J Biol Chem, 2011. **286**(25): p. 22077-89.
201. Li, B., et al., *Increased hedgehog signaling in postnatal kidney results in aberrant activation of nephron developmental programs.* Hum Mol Genet, 2011. **20**(21): p. 4155-66.
202. Marx, A., A. Hoenger, and E. Mandelkow, *Structures of kinesin motor proteins.* Cell Motil Cytoskeleton, 2009. **66**(11): p. 958-66.
203. Friedman, D.S. and R.D. Vale, *Single-molecule analysis of kinesin motility reveals regulation by the cargo-binding tail domain.* Nat Cell Biol, 1999. **1**(5): p. 293-7.
204. Klejnot, M. and F. Kozielski, *Structural insights into human Kif7, a kinesin involved in Hedgehog signalling.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 2): p. 154-9.
205. Liem, K.F., Jr., et al., *Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13377-82.
206. Liu, Y.C., et al., *The PPFIA1-PP2A protein complex promotes trafficking of Kif7 to the ciliary tip and Hedgehog signaling.* Sci Signal, 2014. **7**(355): p. ra117.
207. Tay, S.Y., P.W. Ingham, and S. Roy, *A homologue of the Drosophila kinesin-like protein Costal2 regulates Hedgehog signal transduction in the vertebrate embryo.* Development, 2005. **132**(4): p. 625-34.
208. Endoh-Yamagami, S., et al., *The mammalian Cos2 homolog Kif7 plays an essential role in modulating Hh signal transduction during development.* Curr Biol, 2009. **19**(15): p. 1320-6.
209. Zhulyn, O. and C.C. Hui, *Sufu and Kif7 in limb patterning and development.* Dev Dyn, 2015. **244**(3): p. 468-78.
210. Ye, F., A.R. Nager, and M.V. Nachury, *BBSome trains remove activated GPCRs from cilia by enabling passage through the transition zone.* J Cell Biol, 2018. **217**(5): p. 1847-1868.
211. Cheung, H.O., et al., *The kinesin protein Kif7 is a critical regulator of Gli transcription factors in mammalian hedgehog signaling.* Sci Signal, 2009. **2**(76): p. ra29.
212. Cheung, H.O.-L., et al., *The Kinesin Protein Kif7 Is a Critical Regulator of Gli Transcription Factors in Mammalian Hedgehog Signaling.* 2009. **2**(76): p. ra29-ra29.
213. Niewiadomski, P., et al., *Gli Proteins: Regulation in Development and Cancer.* Cells, 2019. **8**(2).
214. May, S.R., et al., *Loss of the retrograde motor for IFT disrupts localization of Smo to cilia and prevents the expression of both activator and repressor functions of Gli.* Dev Biol, 2005. **287**(2): p. 378-89.
215. Drakopoulou, E., et al., *Non-redundant role for the transcription factor Gli1 at multiple stages of thymocyte development.* Cell Cycle, 2010. **9**(20): p. 4144-52.
216. Park, H.L., et al., *Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation.* Development, 2000. **127**(8): p. 1593-605.
217. Montagnani, V. and B. Stecca, *Role of Protein Kinases in Hedgehog Pathway Control and Implications for Cancer Therapy.* Cancers (Basel), 2019. **11**(4).
218. Bhatia, N., et al., *Gli2 is targeted for ubiquitination and degradation by beta-TrCP ubiquitin ligase.* J Biol Chem, 2006. **281**(28): p. 19320-6.
219. Di Marcotullio, L., et al., *Numb activates the E3 ligase Itch to control Gli1 function through a novel degradation signal.* Oncogene, 2011. **30**(1): p. 65-76.

220. Litingtung, Y. and C. Chiang, *Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3.* Nat Neurosci, 2000. **3**(10): p. 979-85.
221. Zhang, Z., et al., *A proteomic approach identifies SAFB-like transcription modulator (SLTM) as a bidirectional regulator of GLI family zinc finger transcription factors.* J Biol Chem, 2019. **294**(14): p. 5549-5561.
222. Agarwal, N.K., et al., *Active IKKbeta promotes the stability of GLI1 oncogene in diffuse large B-cell lymphoma.* Blood, 2016. **127**(5): p. 605-15.
223. Nye, M.D., et al., *The transcription factor GLI1 interacts with SMAD proteins to modulate transforming growth factor beta-induced gene expression in a p300/CREB-binding protein-associated factor (PCAF)-dependent manner.* J Biol Chem, 2014. **289**(22): p. 15495-506.
224. Lambert, J.P., et al., *Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes.* J Proteomics, 2015. **118**: p. 81-94.
225. Ward, R.J., E. Alvarez-Curto, and G. Milligan, *Using the Flp-In T-Rex system to regulate GPCR expression.* Methods Mol Biol, 2011. **746**: p. 21-37.
226. Schmidt, T.G. and A. Skerra, *The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins.* Nat Protoc, 2007. **2**(6): p. 1528-35.
227. Choi, H., et al., *SAINT: probabilistic scoring of affinity purification-mass spectrometry data.* Nat Methods, 2011. **8**(1): p. 70-3.
228. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification-mass spectrometry data.* Nat Methods, 2013. **10**(8): p. 730-6.
229. Shin, C.J., et al., *Protein-protein interaction as a predictor of subcellular location.* BMC Syst Biol, 2009. **3**: p. 28.
230. Cosson, P., et al., *Dynamic transport of SNARE proteins in the Golgi apparatus.* Proc Natl Acad Sci U S A, 2005. **102**(41): p. 14647-52.
231. Beurel, E., S.F. Grieco, and R.S. Jope, *Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases.* Pharmacol Ther, 2015. **148**: p. 114-31.
232. Gangoda, L., et al., *Extracellular vesicles including exosomes are mediators of signal transduction: are they protective or pathogenic?* Proteomics, 2015. **15**(2-3): p. 260-71.
233. Kondelin, J., et al., *Comprehensive evaluation of coding region point mutations in microsatellite-unstable colorectal cancer.* EMBO Mol Med, 2018. **10**(9).
234. Keskitalo, S., et al., *Novel TMEM173 Mutation and the Role of Disease Modifying Alleles.* Front Immunol, 2019. **10**: p. 2770.
235. Koc, E.C., et al., *Identification and characterization of CHCHD1, AURKAIP1, and CRIF1 as new members of the mammalian mitochondrial ribosome.* Front Physiol, 2013. **4**: p. 183.
236. Neupert, W. and J.M. Herrmann, *Translocation of Proteins into Mitochondria.* 2007. **76**(1): p. 723-749.
237. Voos, W. and K. Röttgers, *Molecular chaperones as essential mediators of mitochondrial biogenesis.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2002. **1592**(1): p. 51-62.
238. Goswami, A.V., B. Chittoor, and P. D'Silva, *Understanding the functional interplay between mammalian mitochondrial Hsp70 chaperone machine components.* J Biol Chem, 2010. **285**(25): p. 19472-82.
239. Srivastava, S., et al., *Regulation of mitochondrial protein import by the nucleotide exchange factors GrpEL1 and GrpEL2 in human cells.* J Biol Chem, 2017. **292**(44): p. 18075-18090.
240. Varjosalo, M., et al., *Application of active and kinase-deficient kinome collection for identification of kinases regulating hedgehog signaling.* Cell, 2008. **133**(3): p. 537-48.
241. Onishi H, K.M., *Hedgehog signaling pathway as a therapeutic target in various types of cancer.* Cancer Sci., 2011. **102**(10): p. 1756-60.
242. Varjosalo M, L.S., Taipale J., *Divergence of hedgehog signal transduction mechanism between Drosophila and mammals.* Dev Cell, 2006. **10**(2): p. 177-86.
243. Pusapati, G.V., et al., *CRISPR Screens Uncover Genes that Regulate Target Cell Sensitivity to the Morphogen Sonic Hedgehog.* Dev Cell, 2018. **44**(1): p. 113-129.e8.

244. Goto, H., A. Inoko, and M. Inagaki, *Cell cycle progression by the repression of primary cilia formation in proliferating cells.* Cell Mol Life Sci, 2013. **70**(20): p. 3893-905.
245. Yang, H., et al., *Opposite regulation of Wnt/β-catenin and Shh signaling pathways by Rack1 controls mammalian cerebellar development.* 2019. **116**(10): p. 4661-4670.
246. Ding, M. and X. Wang, *Antagonism between Hedgehog and Wnt signaling pathways regulates tumorigenicity.* Oncol Lett, 2017. **14**(6): p. 6327-6333.
247. Caparros-Martin, J.A., et al., *Specific variants in WDR35 cause a distinctive form of Ellis-van Creveld syndrome by disrupting the recruitment of the EvC complex and SMO into the cilium.* Hum Mol Genet, 2015. **24**(14): p. 4126-37.
248. Gupta, G.D., et al., *A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface.* Cell, 2015. **163**(6): p. 1484-99.
249. D'Amico, D., et al., *Non-canonical Hedgehog/AMPK-Mediated Control of Polyamine Metabolism Supports Neuronal and Medulloblastoma Cell Growth.* Dev Cell, 2015. **35**(1): p. 21-35.
250. Liu, X., et al., *An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations.* Nat Commun, 2018. **9**(1): p. 1188.
251. Youn, J.Y., et al., *High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies.* Mol Cell, 2018. **69**(3): p. 517-532.e11.
252. Lauth, M. and R. Toftgard, *Non-canonical activation of GLI transcription factors: implications for targeted anti-cancer therapy.* Cell Cycle, 2007. **6**(20): p. 2458-63.
253. Satir, P., L.B. Pedersen, and S.T. Christensen, *The primary cilium at a glance.* 2010. **123**(4): p. 499-503.
254. Caspary, T., C.E. Larkins, and K.V. Anderson, *The graded response to Sonic Hedgehog depends on cilia architecture.* Dev Cell, 2007. **12**(5): p. 767-78.

ISBN 978-951-51-6010-2 (PRINT)
ISBN 978-951-51-6011-9 (ONLINE)
ISSN 2342-3161 (PRINT)
ISSN 2342-317X (ONLINE)
http://ethesis.helsinki.fi

HELSINKI 2020